

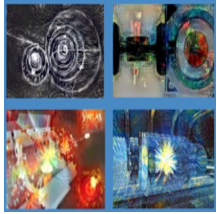
Graph Networks in High Energy Physics

*Fifth Machine Learning in High
Energy Physics Summer School
1-10 July 2019*

Jean-Roch Vlimant, jvlimant@caltech.edu



Outline



I. Physics at the LHC

II. The case for Machine Learning

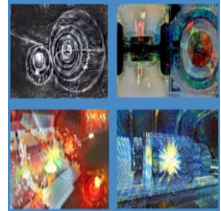
III. Collider Data Representation

IV. Graph Networks & applications

Covering the following

- *Tracking with GNN*
- *Learning irregular geometry with Distance-weighted GNN*
- *Pile-Up mitigation with GGCNN*
- *Jet identification with Interaction network*

Relevant work not covered: [icecube](#), [MPJet](#), [CloudJet](#), [stopGNN](#), [showerGAN](#)

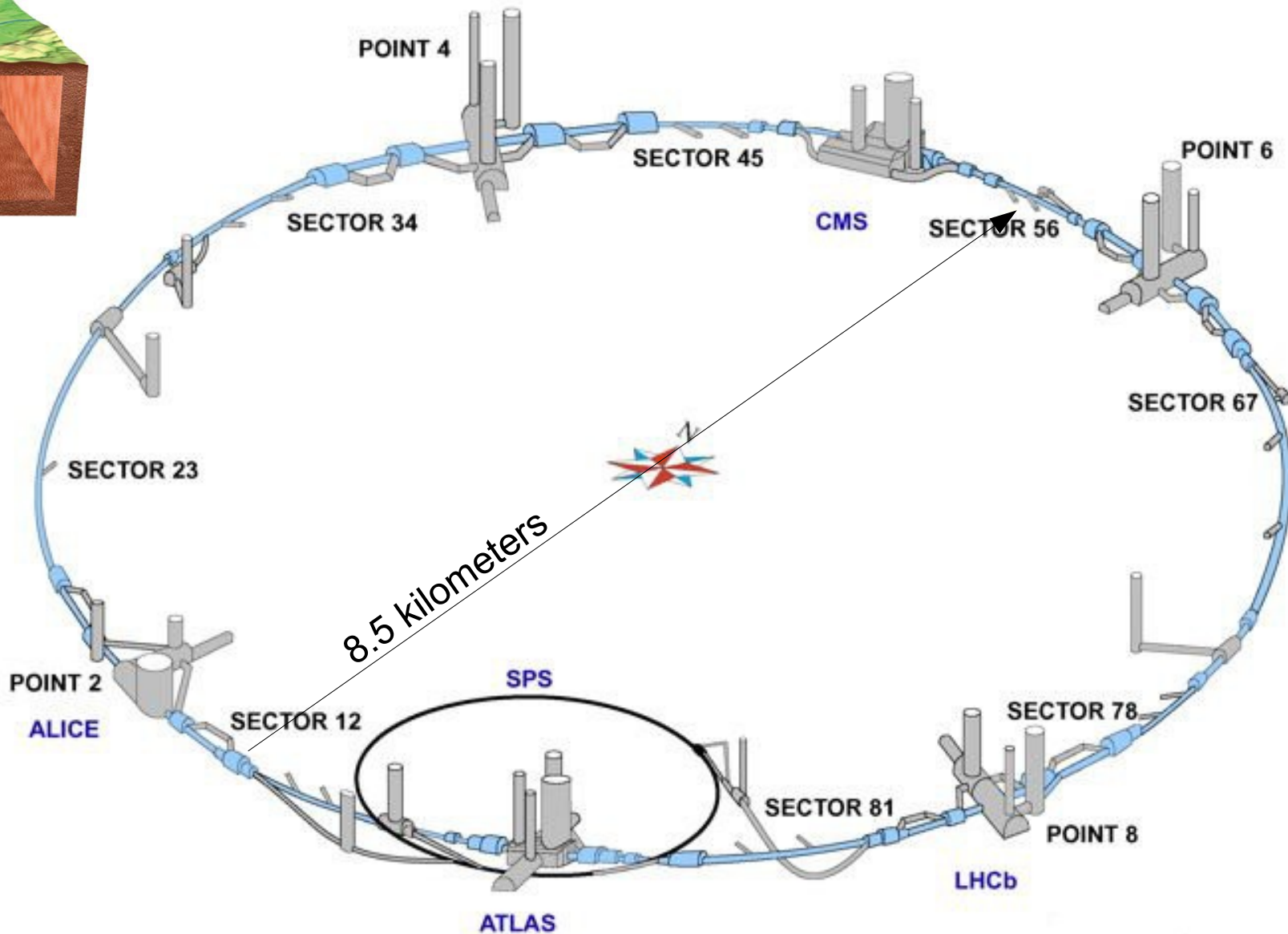
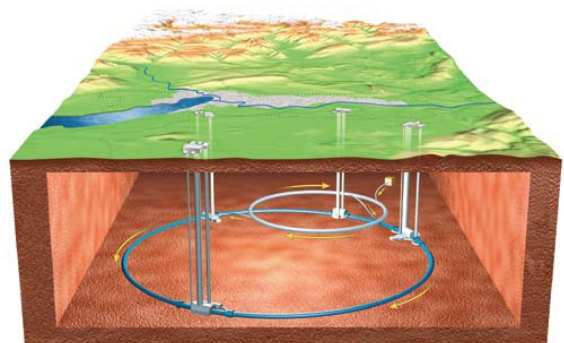
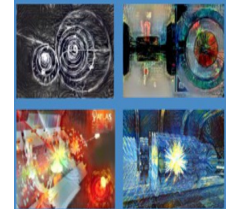


High Energy Physics Endeavor

In a nutshell

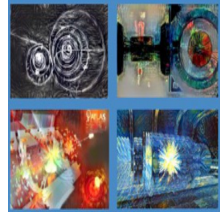


The Large Hadron Collider

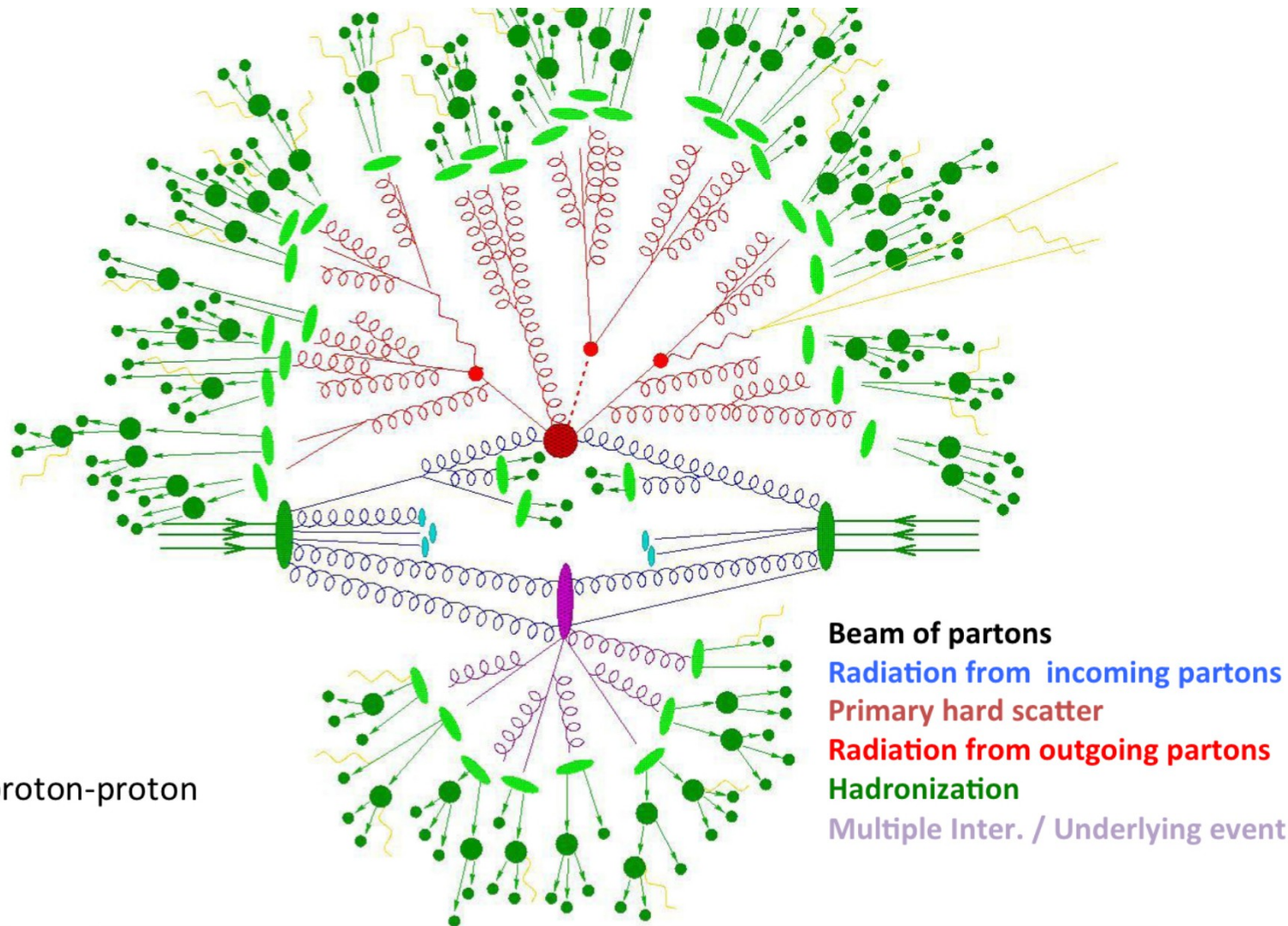




Colliding Hadrons



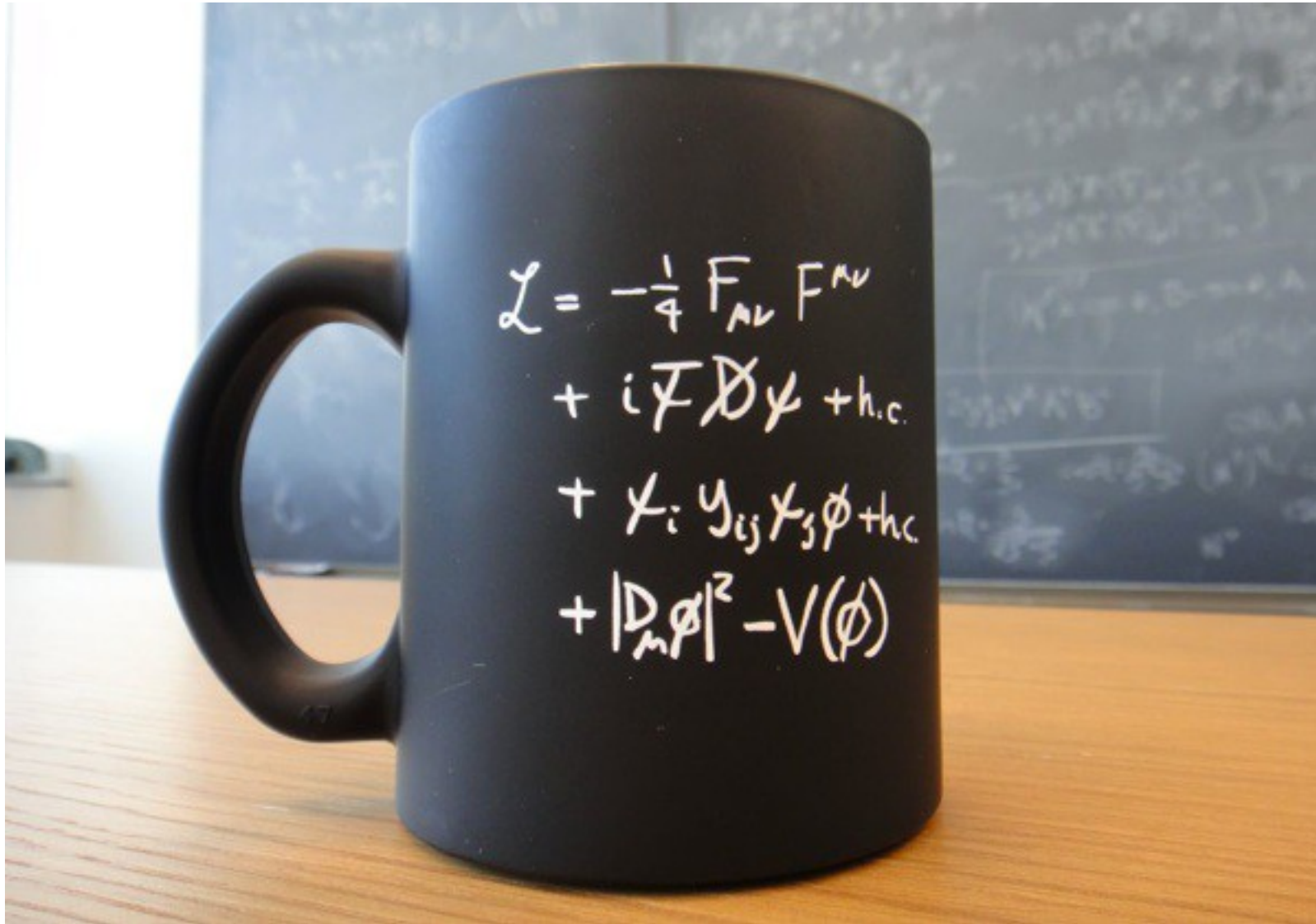
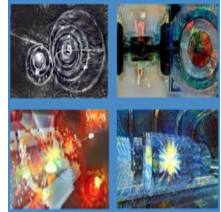
Typical proton-proton collision



Probing fundamental laws of physics as large spectrum of particles (known and unknown) can be produced



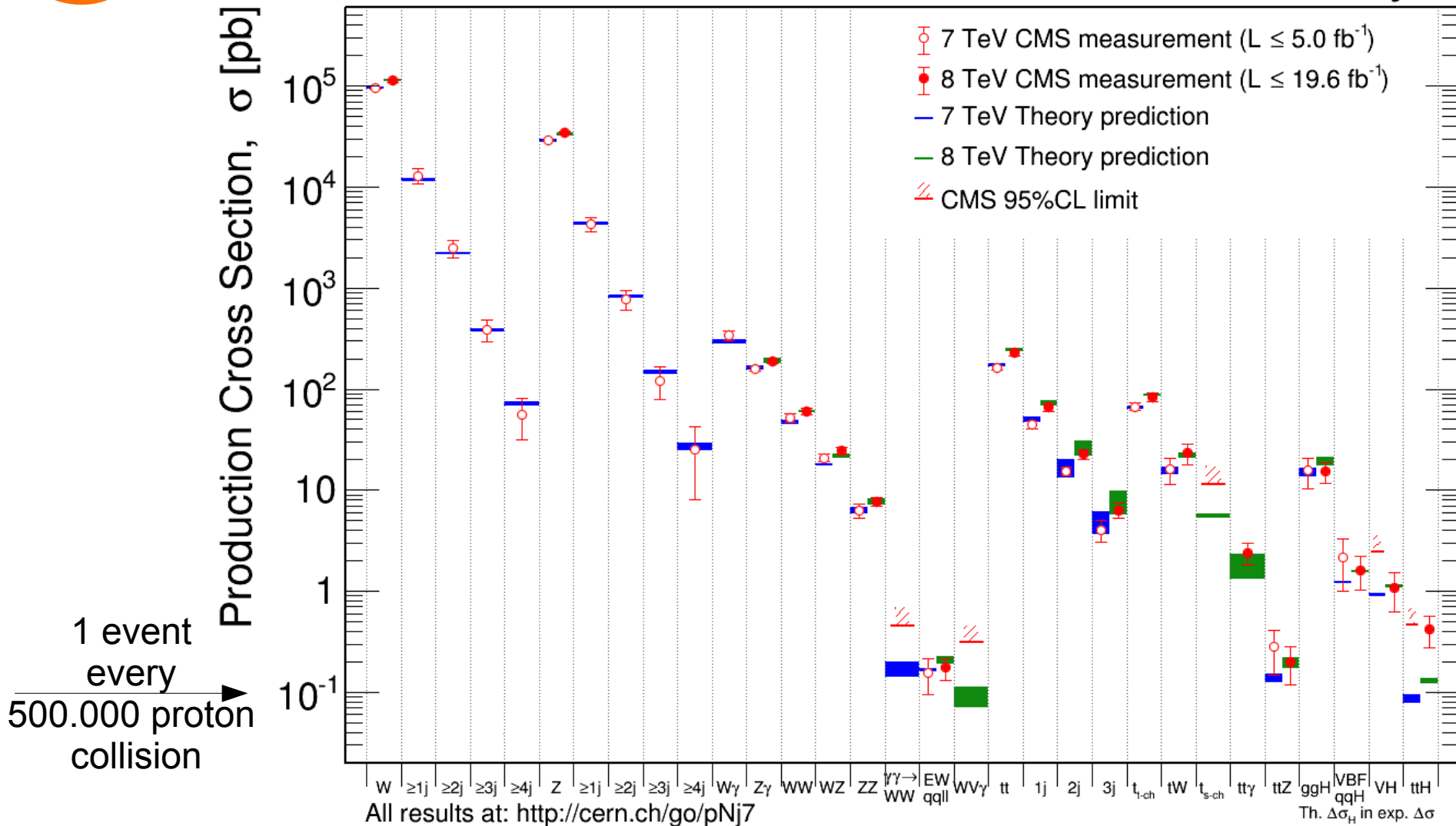
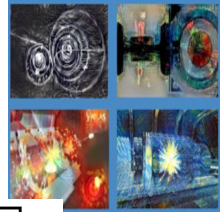
The Standard Model



Well demonstrated effective model
We can predict most of the observations
We can use a large amount of simulation



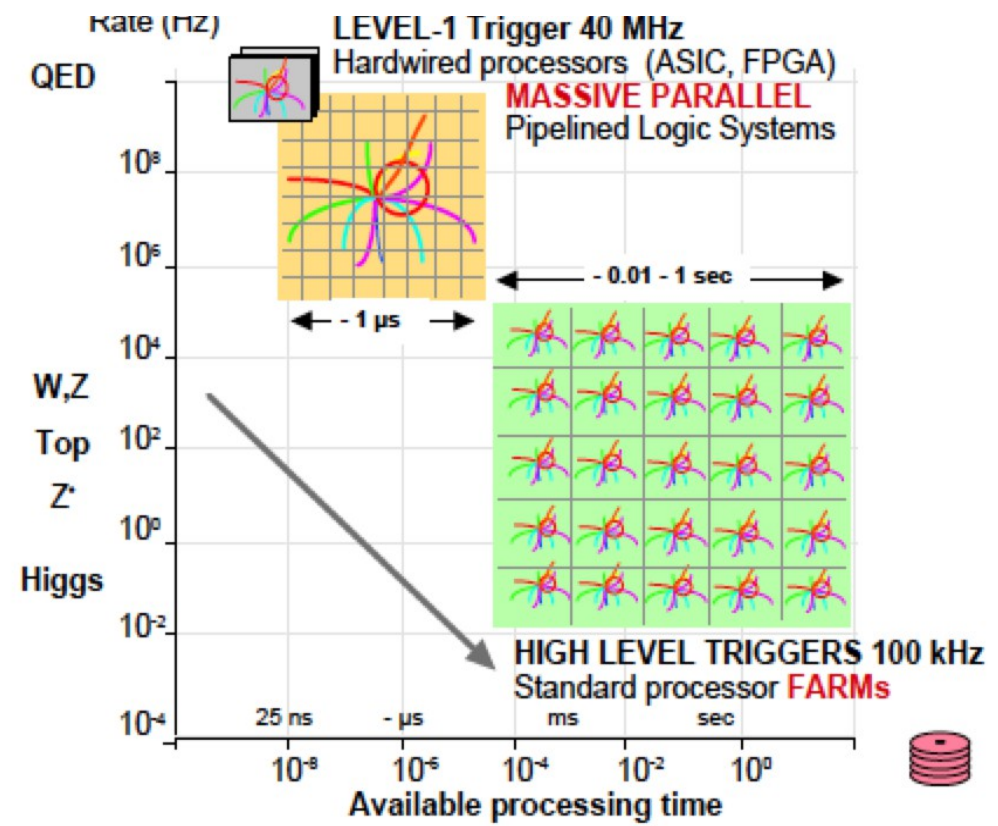
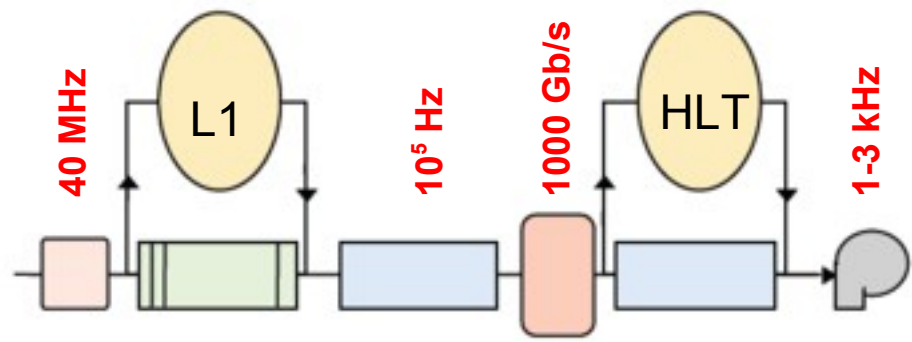
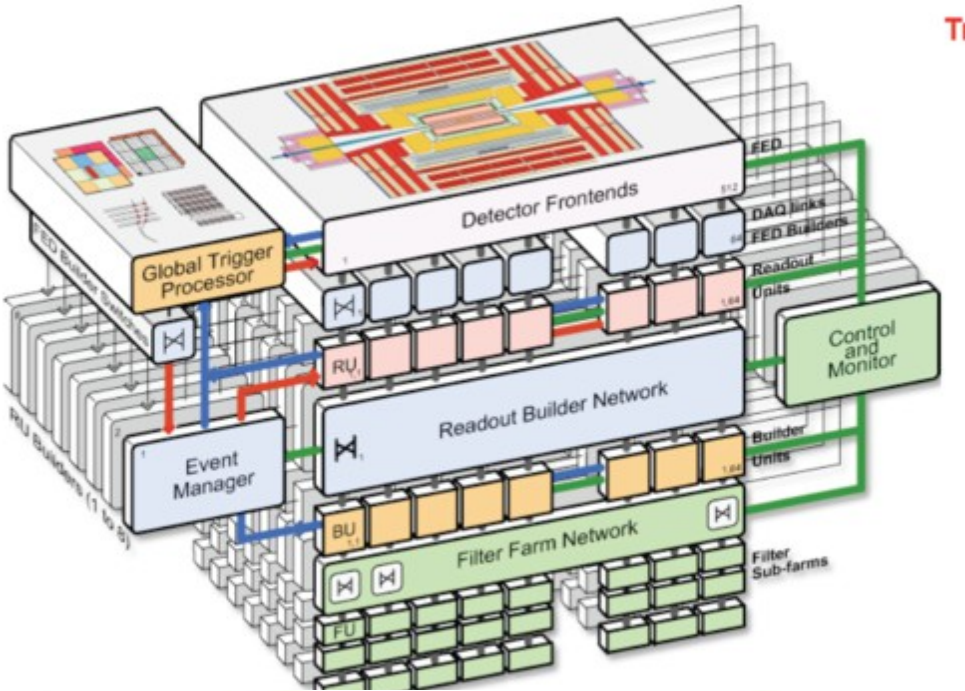
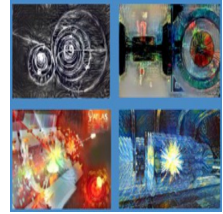
Size Of The Challenge



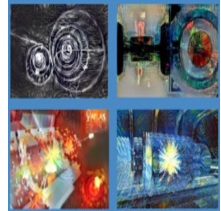
Low probability of producing exotic and interesting signals.
Observe rare events from a large amount of data.



Event Filtering



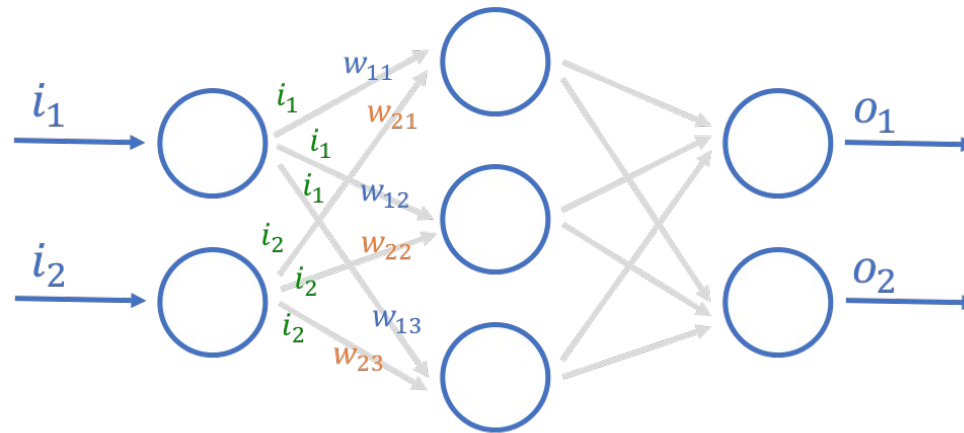
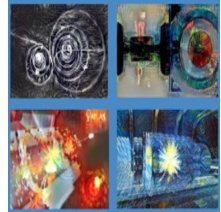
Ultra fast decision to keep the relevant data.
In hardware and software.



The Case for Machine Learning



Operation Vectorisation



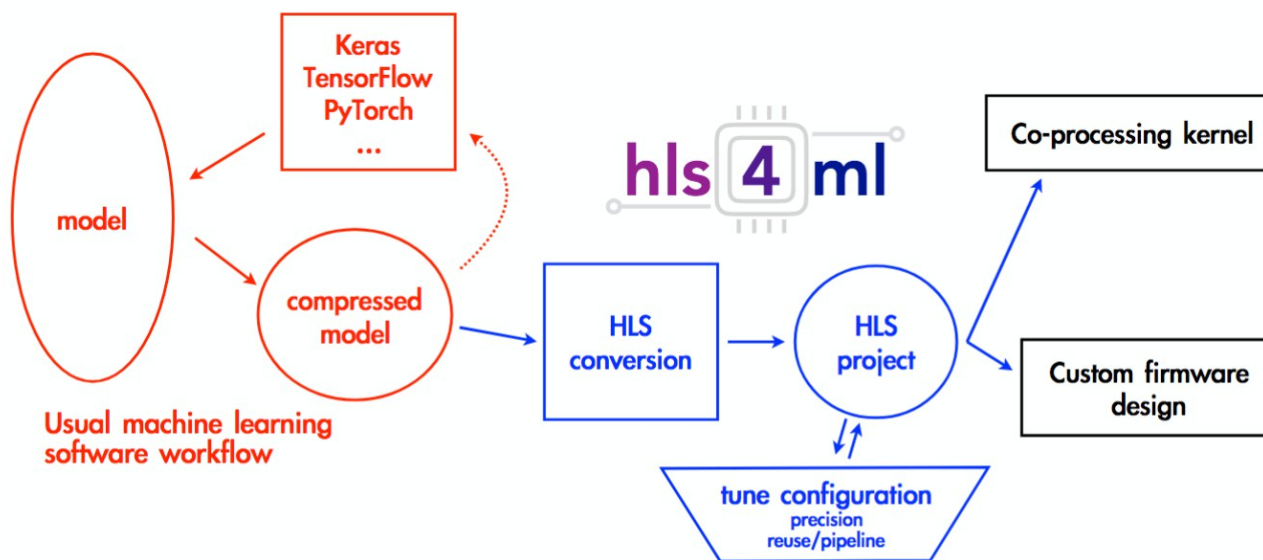
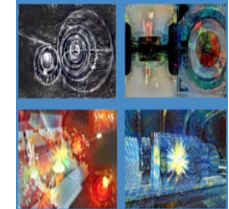
ANN \equiv matrix operations \equiv parallelizable

$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

Computation of prediction from artificial neural network model can be **vectorized to a large extend.**



Hyper-Fast Prediction



Synthesis FPGA firmware
from trained ANN

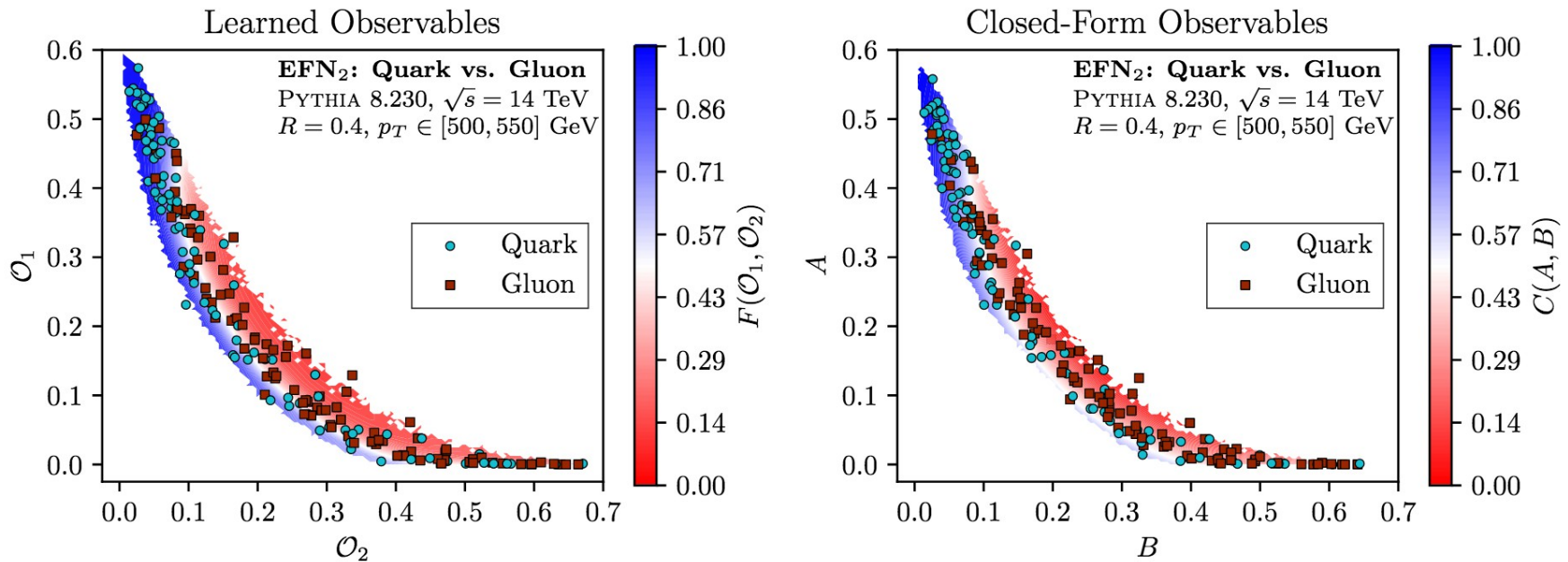
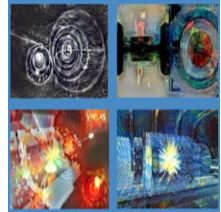
<https://hls-fpga-machine-learning.github.io/hls4ml/>

J. Duarte et al. <https://arxiv.org/abs/1804.06913>

Prediction from artificial neural network model can be
done on FPGA, GPU, TPU, ...



Physics Knowledge

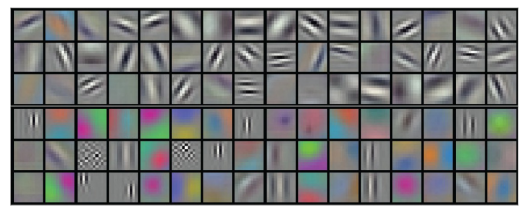
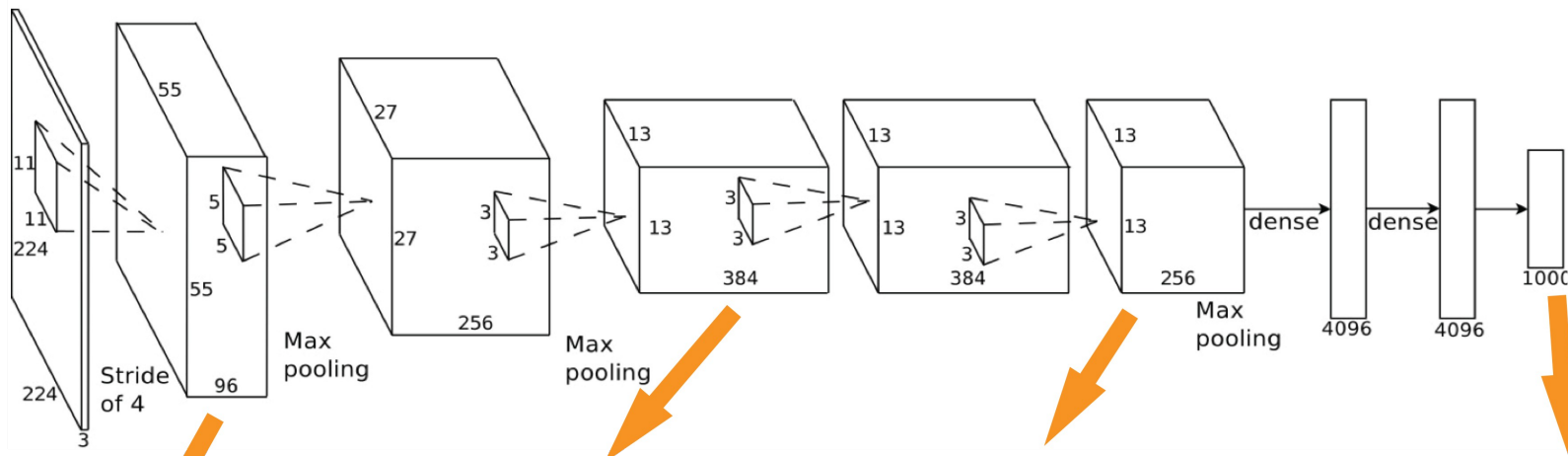
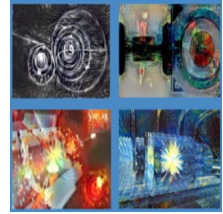


P. Komiske, E. Metodiev, J. Thaler, <https://arxiv.org/abs/1810.05165>

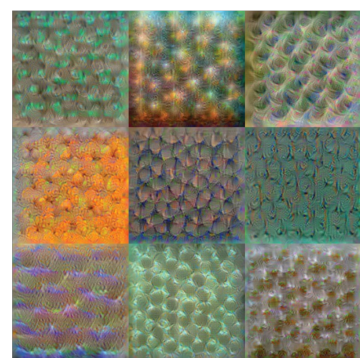
Machine Learning can help understand Physics.
We can make better models with Physics.



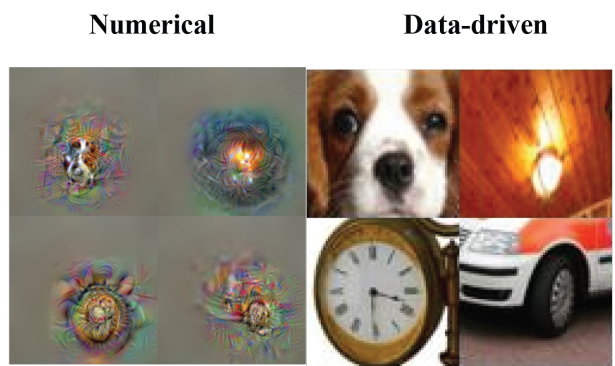
Learning from Complexity



Conv 1: Edge+Blob



Conv 3: Texture



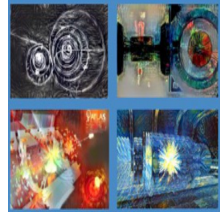
Conv 5: Object Parts



Fc8: Object Classes

“Simple” machine learning model can **extract information from complex dataset.**

More classical algorithm counter part may take **years of development.**



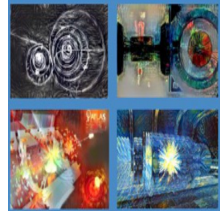
Take home message :

Measure rare and exotic processes from orders of magnitude larger backgrounds.

The Standard Model predicts with precision what to expect from many processes.

Reconstruct, identify and reject large amount of event within resource constraints.

Machine learning is a path forward in a resource constrained environment.

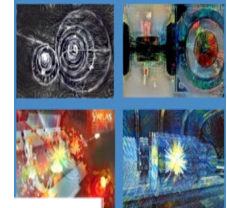


High Energy Physics Data Representation

With bias on CMS

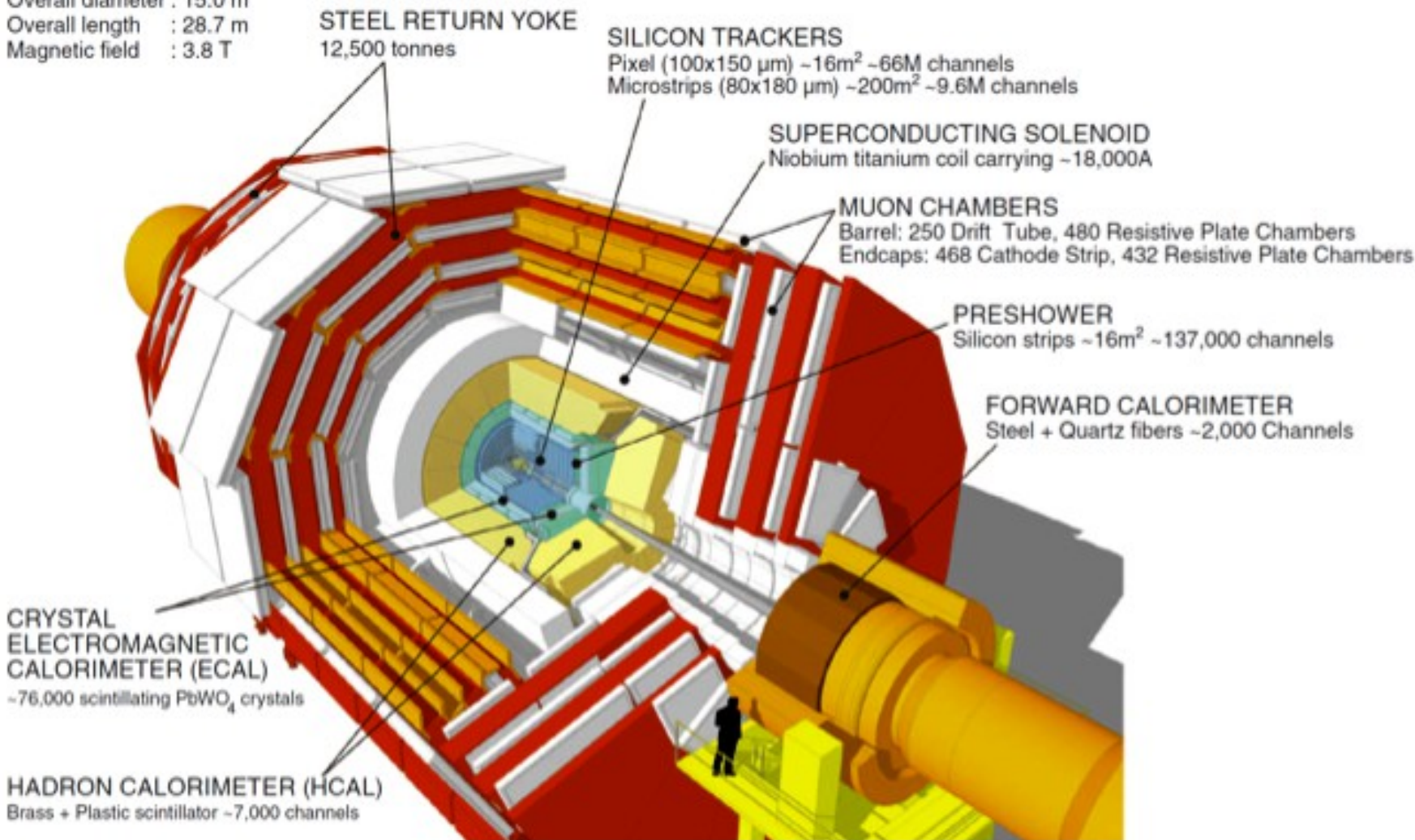


CMS Detector



CMS DETECTOR

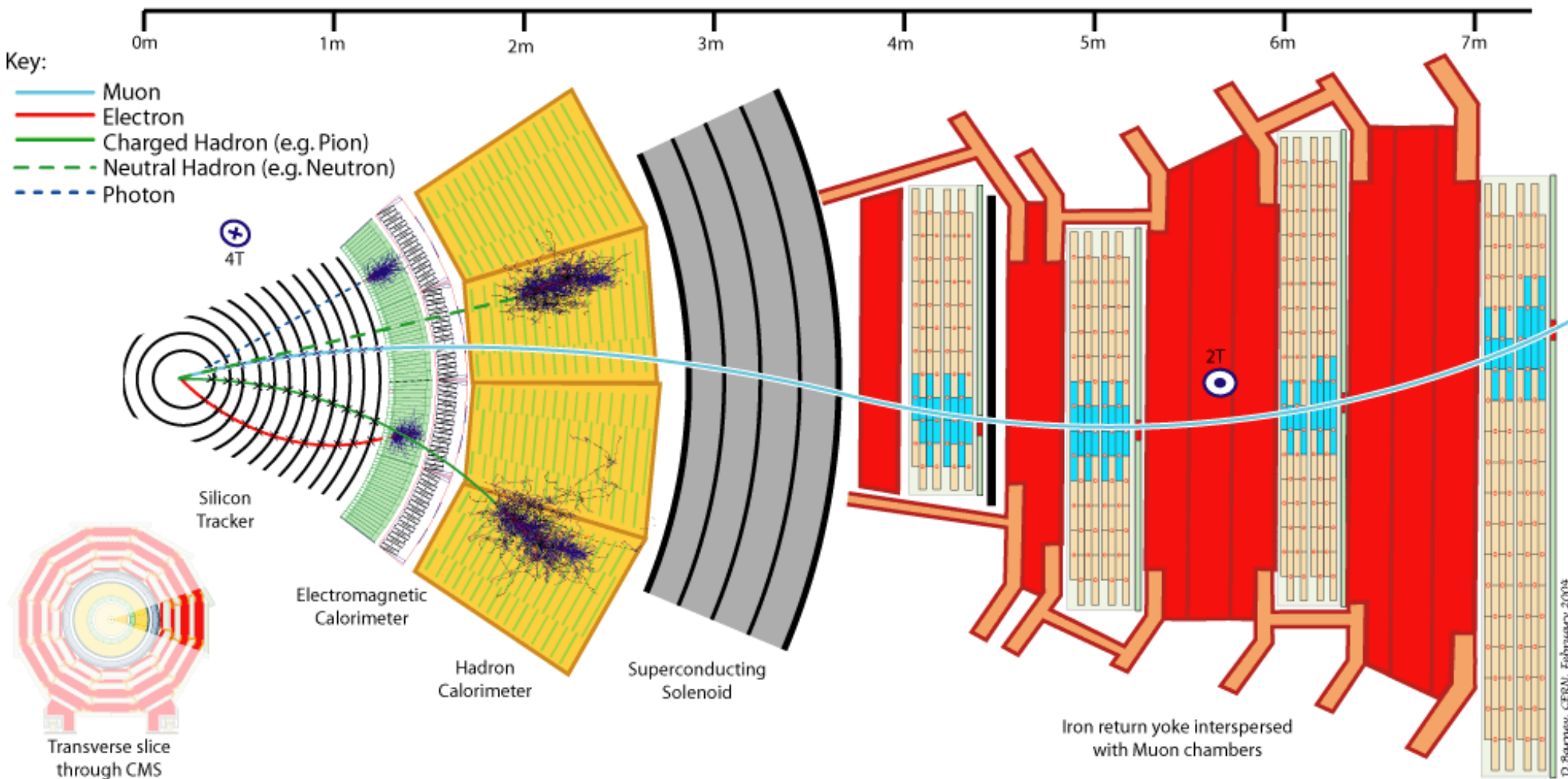
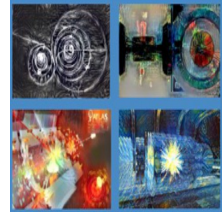
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T



Heterogenous detector, with complex geometry.
About 100M channels in the read-out.



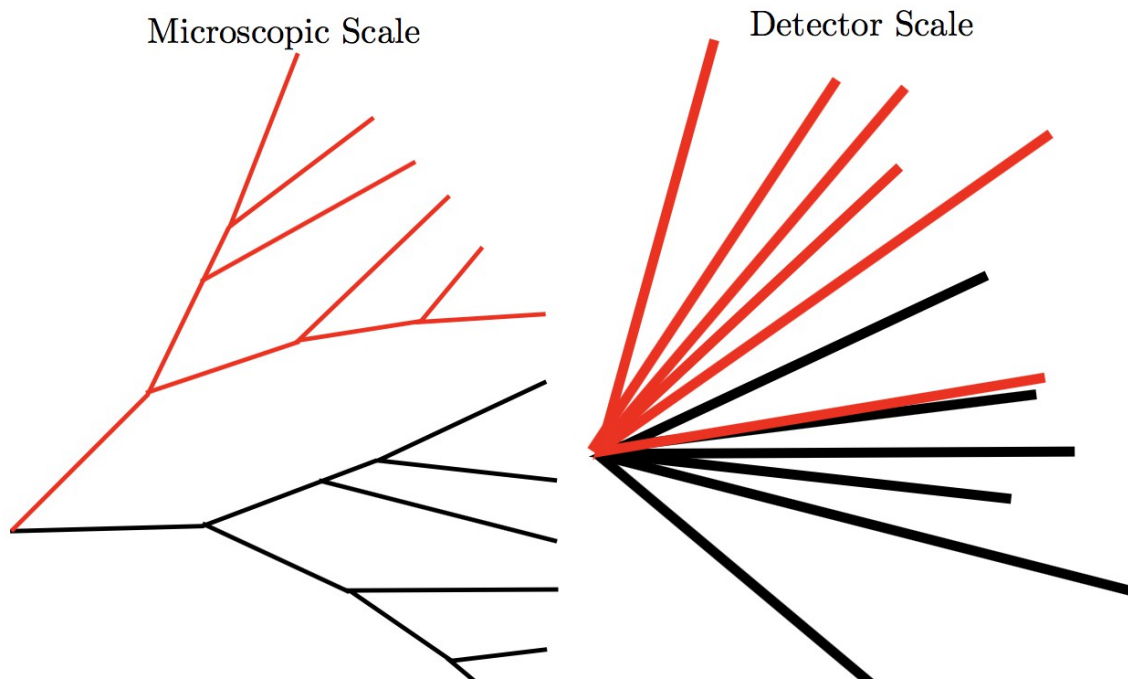
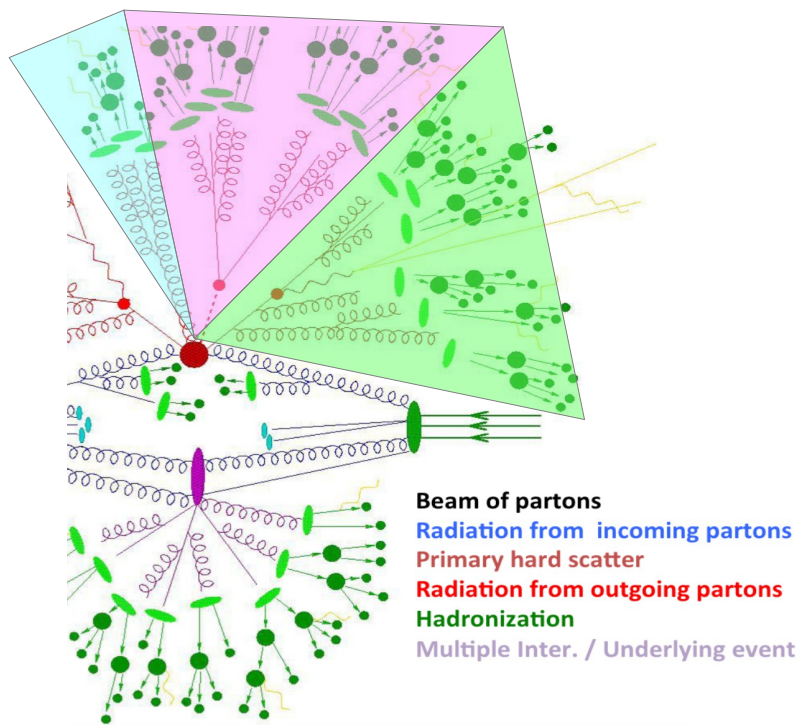
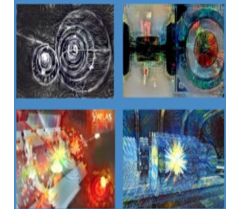
A Journey Through Matter



Particles leave hints of their passage in sub-detectors. Specific (but overlapping) pattern for each particle type.



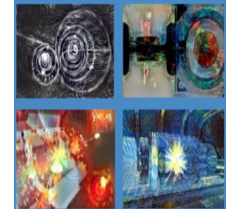
What is a Jet



Quark&gluons hadronize as they propagate.
Any particle decaying in quark/gluons will result in a “jet” of particles in the direction of the original particle.
Ambiguities on the original particle gets worse in boosted systems.



From RAW to High Level data



Detector Data

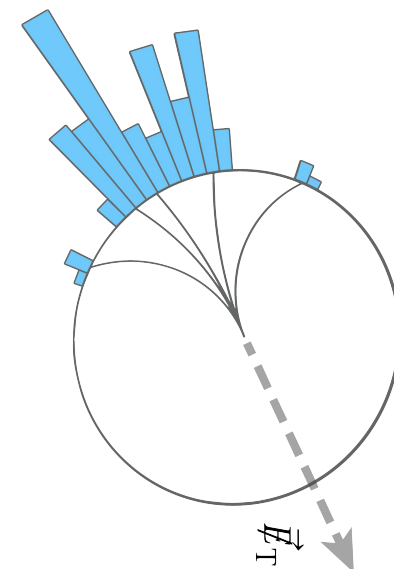
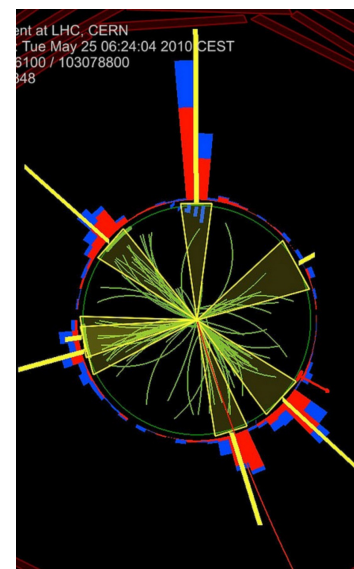
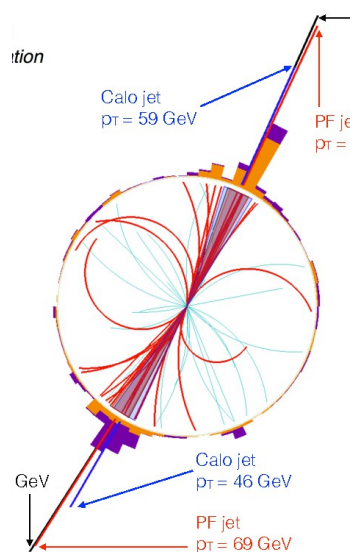
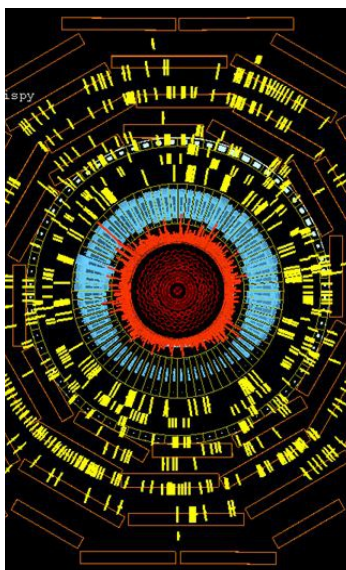
Local reconstruction

Particle representation

Jet Clustering

High level features

```
1A 16 76 C5
6C FF C2 E5
ADC1 B3 3B
36 36 E4 EE
97 13 16 FA
1B 68 FF E8
6A 41 C1 1A
E8 E4 CD 99
1A 16 76 C5
6C FF C2 E5
ADC1 B3 3B
```



Event Processing

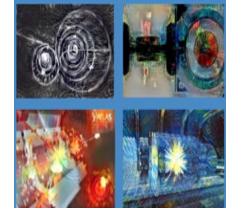
Dimensionality reduction

Globalization of information

The reconstruction of an event goes from the digital signal of the individual sub-detector to a sequence of particles, jets, and high-level features

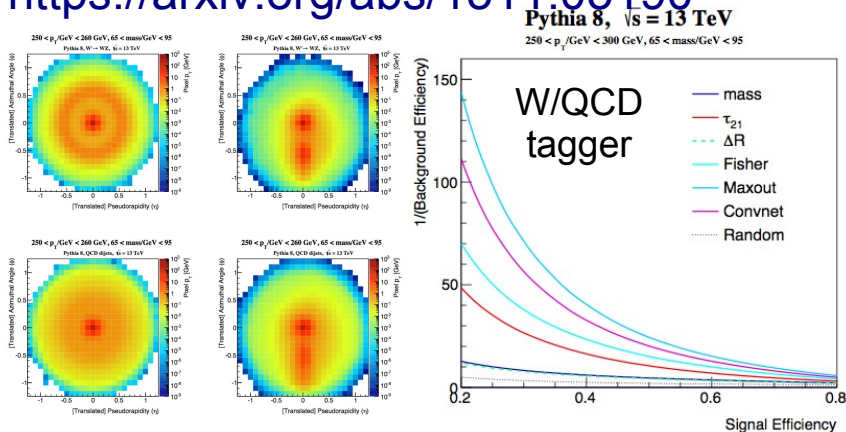


Images and Sequences



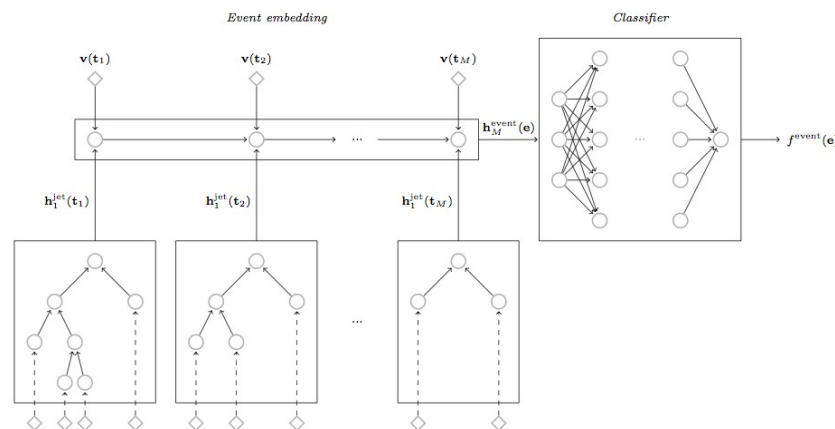
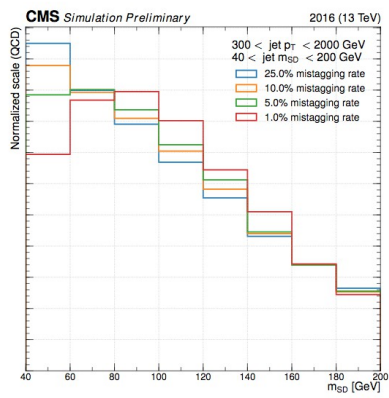
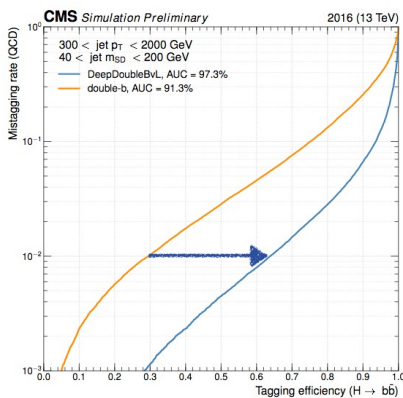
Jet Imaging

<https://arxiv.org/abs/1511.05190>



DP-2018/033 DEEP DOUBLE-B TAGGER

- Large performance gain over BDT
- Default algorithm still “learns” the mass \Rightarrow mass sculpting

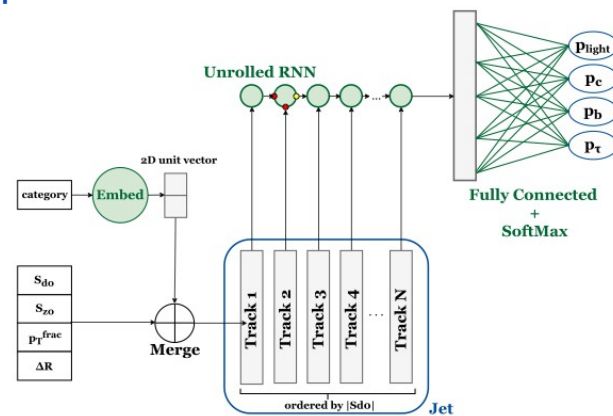


QCD-Aware Recursive Neural Networks for Jet Physics.

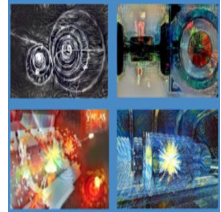
<https://arxiv.org/abs/1702.00748>

B-Jet with Recurrent Neural Networks

<http://cds.cern.ch/record/2255226>



Possible loss of information with image representation.
 Choice on ordering with sequence representation.

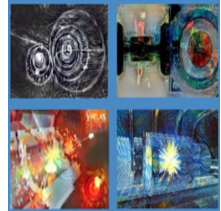


Take home message :

An event is a snapshot of the hints left by thousands to tens of thousands of particles.

Particle identification is mostly a pattern recognition task.

Graph-like data representation seems natural at many level.

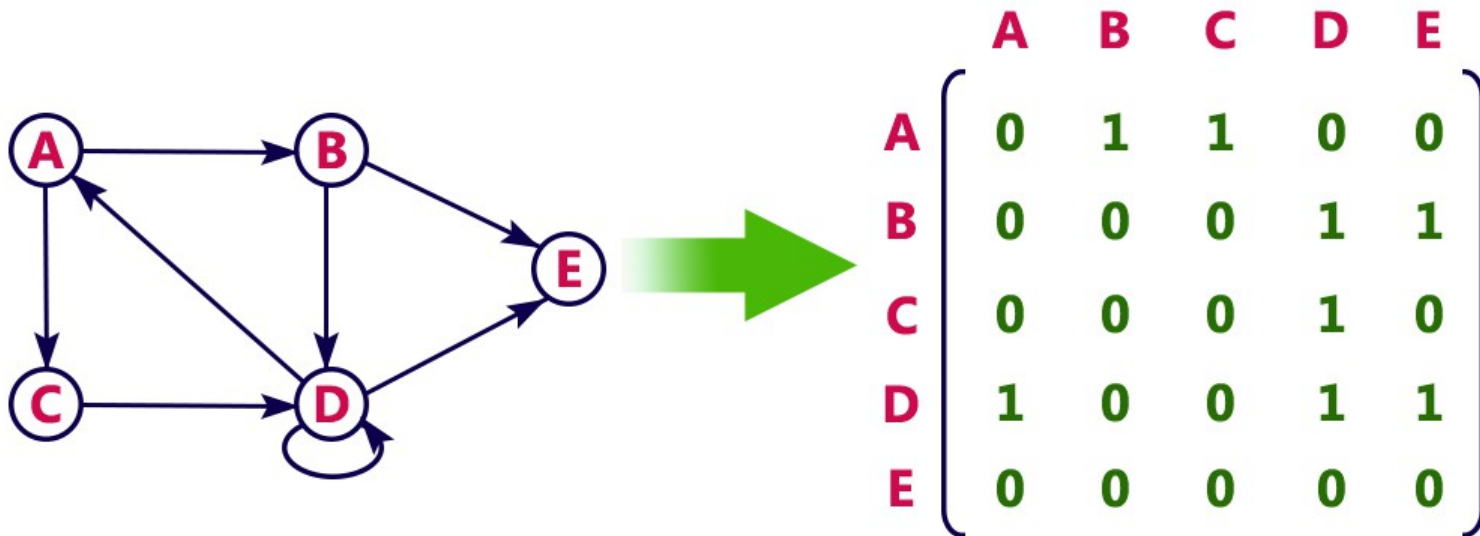
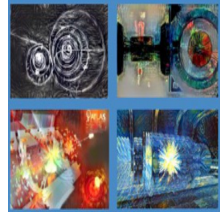


Graph Network applications

Further than image and sequences



Forewords on Graph



http://btechsmartclass.com/data_structures/graph-representations.html

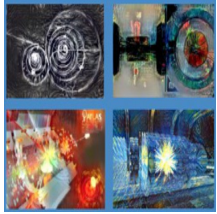
A graph is composed of

- **Nodes** that can be represented as a vector.
- **Edges** that can be represented with the adjacency matrix.

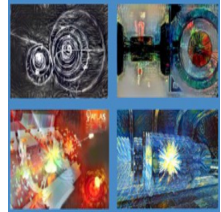
- Flowing of information using matrix operations.
- With machine learning on graphs, edges and nodes might acquire internal representations.



Overview



- **Charged particle tracking**
Connecting the sparse hits left by particles along trajectory
- **Calorimeter reconstruction**
Assemble pattern of energy depositions
- **Pile-up mitigation**
Reducing the impact of concurrent proton-proton interactions
- **Jet Identification**
Unveil the origin of collimated spray of particles



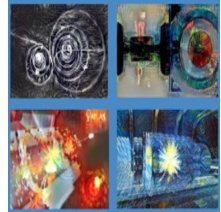
Charged Particle Tracking

Clustering sparsely measured hits into trajectory of charged particles.

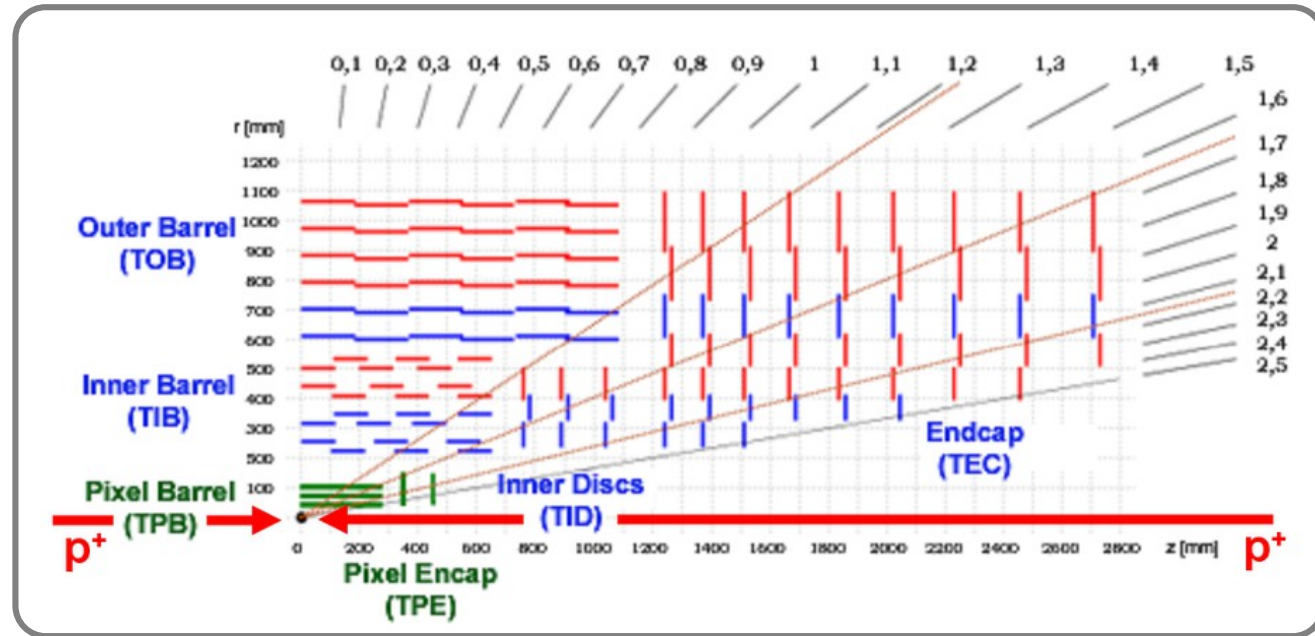
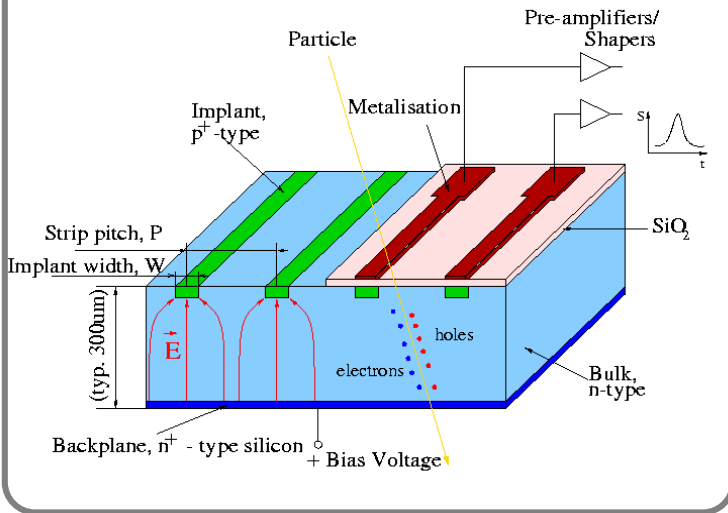
With Steven Farrell, Paolo Calafiura, Mayur Mudigonda, Prabhat, Dustin Anderson, Stephan Zheng, Josh Bendavid, Maria Spiropulu, Giuseppe Cerati, Lindsey Gray, Jim Kowalkowski, Panagiotis Spentzouris, Aristeidis Tsaris, Xiangyang Yu
<https://arxiv.org/abs/1810.06111>



Tracker – 101



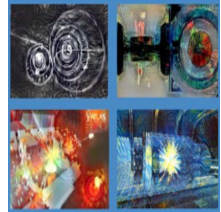
Principles of operation



- Ionizing particle leaves a signal in silicon detector
- Sensitive cells with different shape
 - squares : pixel
 - rectangle : strips
- Organized in hermetic layers of modules

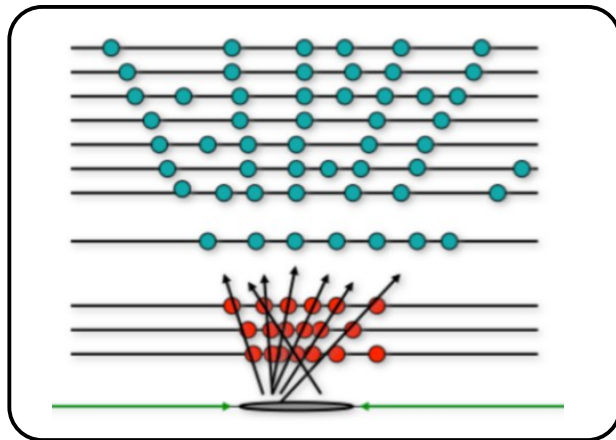


Particle Tracking – 101

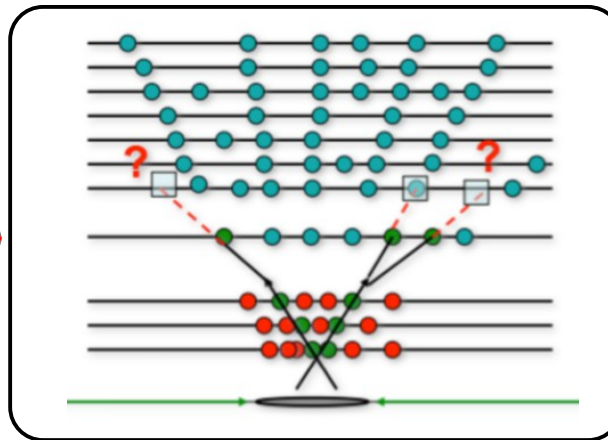


- Particle trajectory bended in a solenoidal magnetic field
 - Curvature is a proxy to momentum
- **Thousands of sparse hits**
- Hits pollution from low momentum, secondary particles

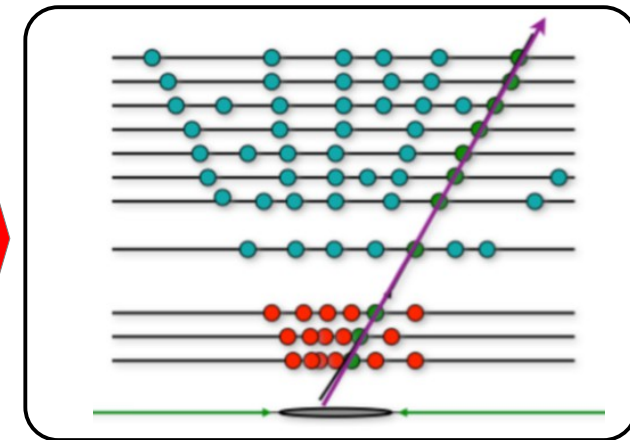
Seeding



Combinatorial Kalman Filter



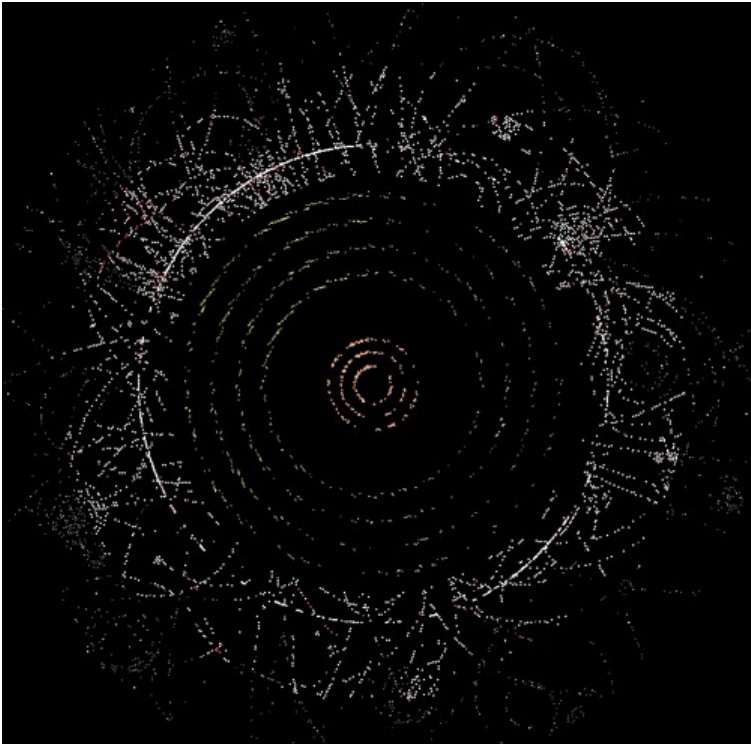
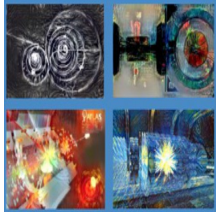
Fitting with Kalman Filter



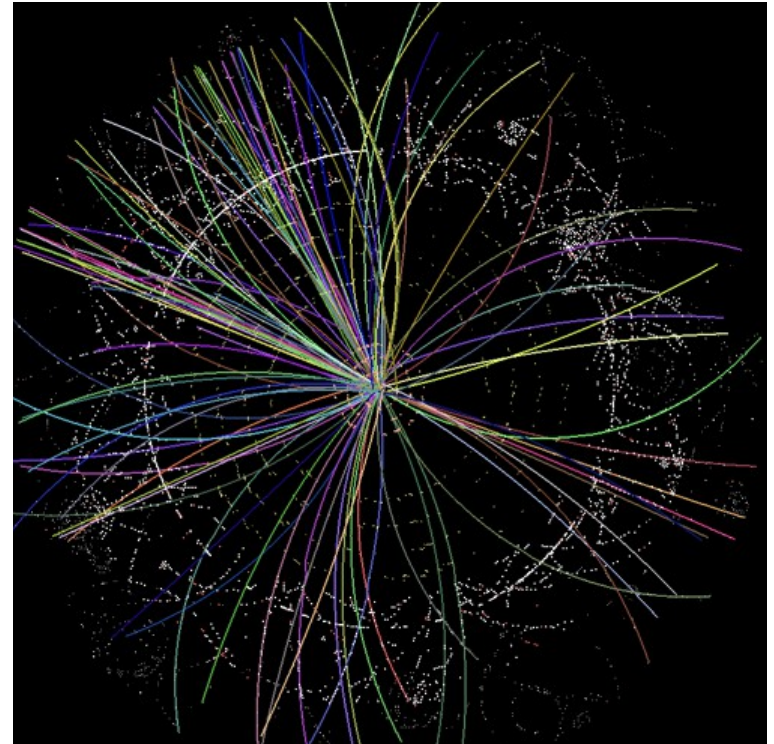
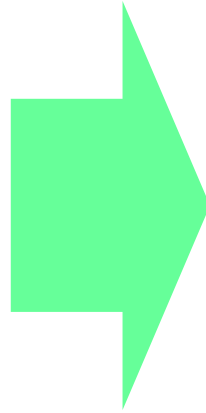
- **Explosion of hit combinatorics** in both seeding and stepping pattern recognition
- **Highly computing consuming task** in extracting physics content from LHC data



Name of the Game



From hits ...



... to trajectory & parameters



Other ML Solutions

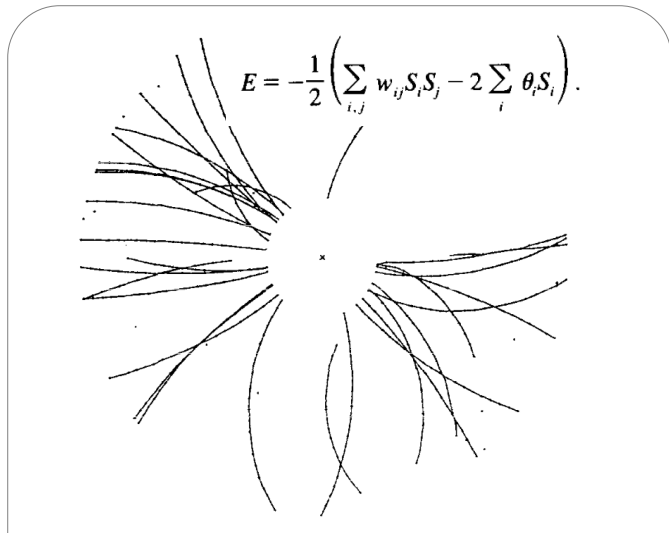
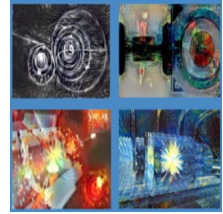
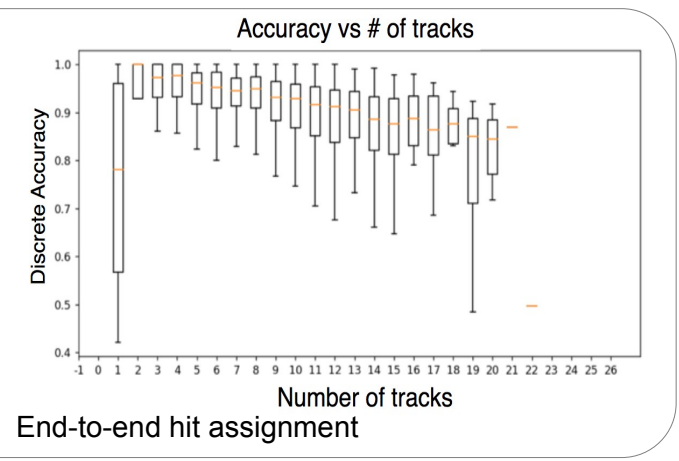
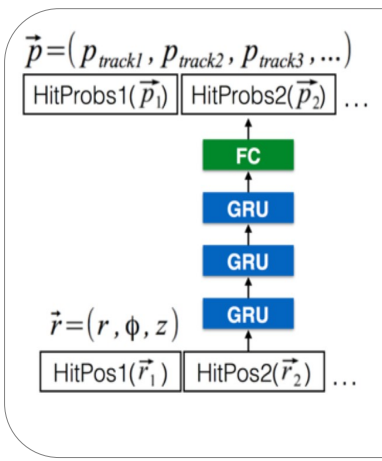
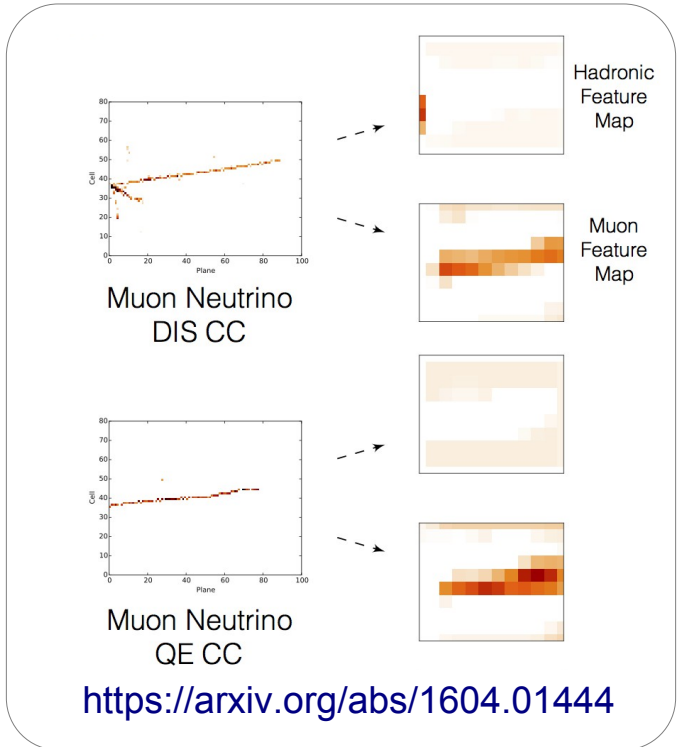
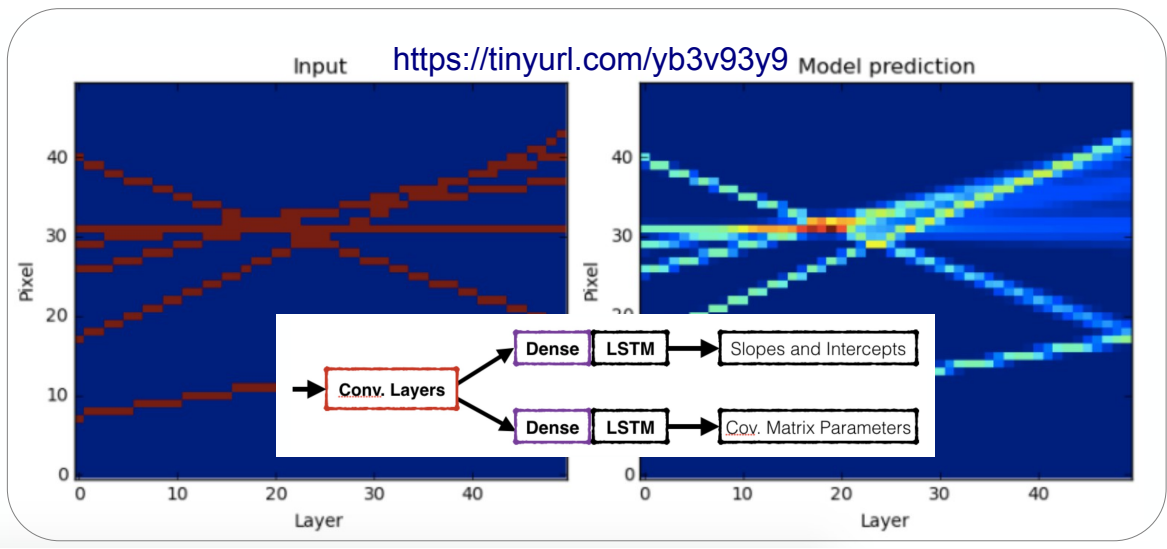


Fig. 4. Tracks in the ALEPH TPC reconstructed with a Hopfield net [13]. <https://tinyurl.com/y9swquw6>



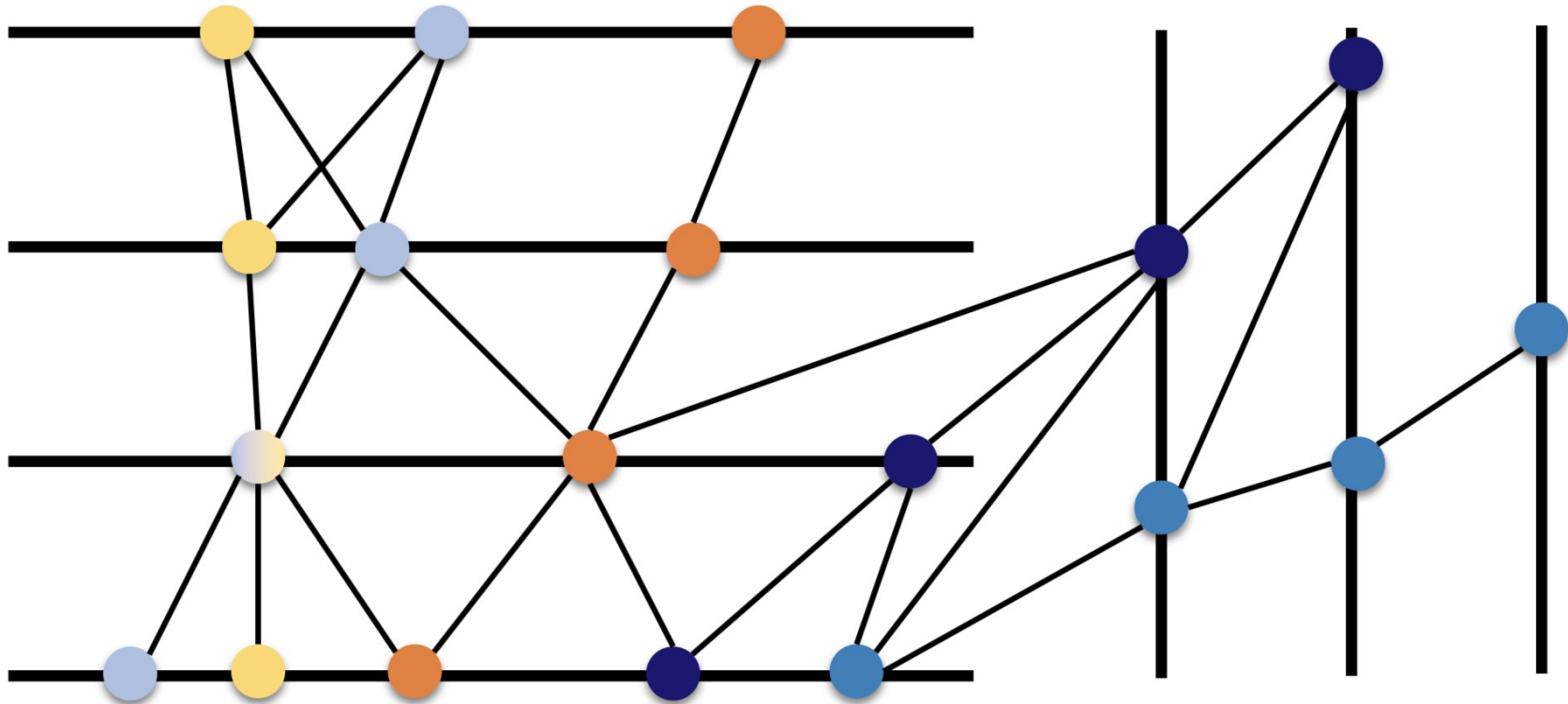
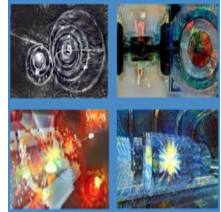
End-to-end hit assignment



<https://arxiv.org/abs/1604.01444>



Tracker Hit Graph

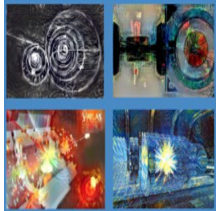


Graph construction

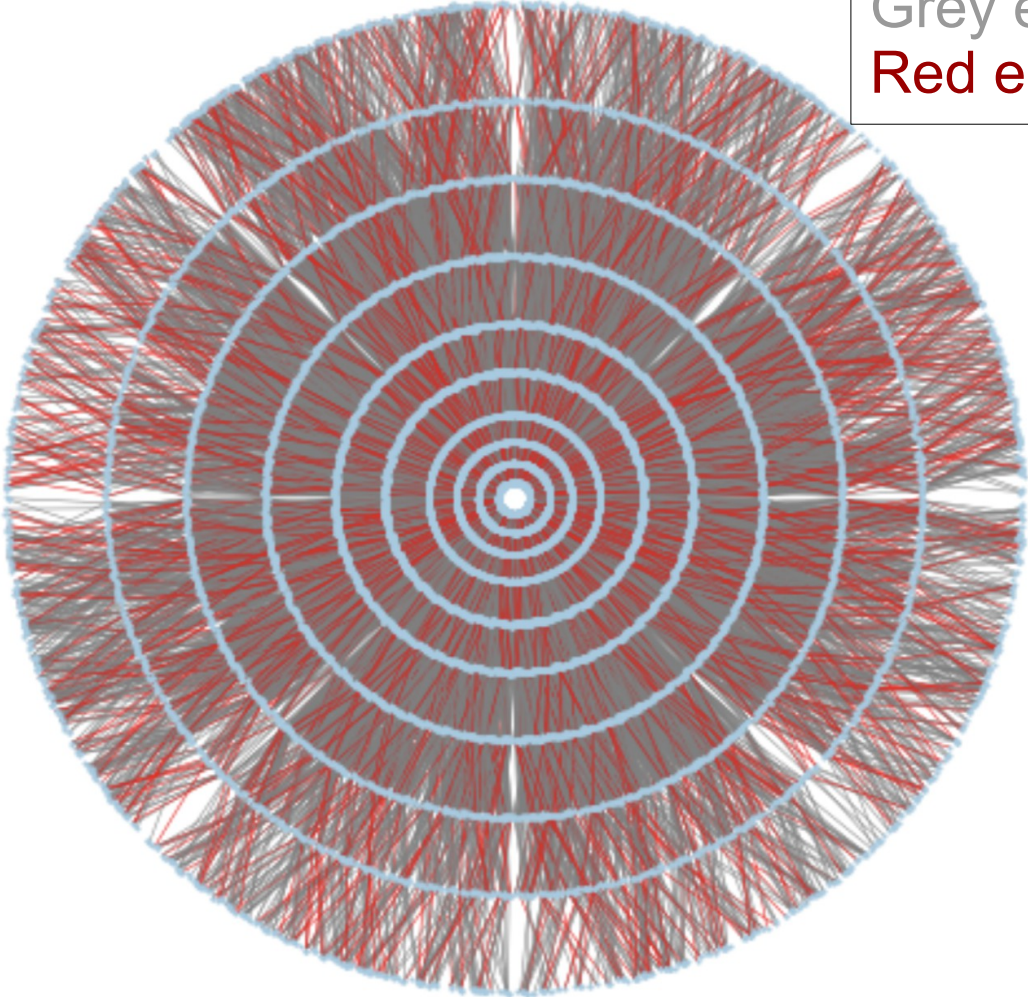
- One tracker hit \equiv one node
- Sparse edges constructed from geometrical consideration
- ➔ Edge classification \equiv reconstructing the trajectory of particles



Edge Classification

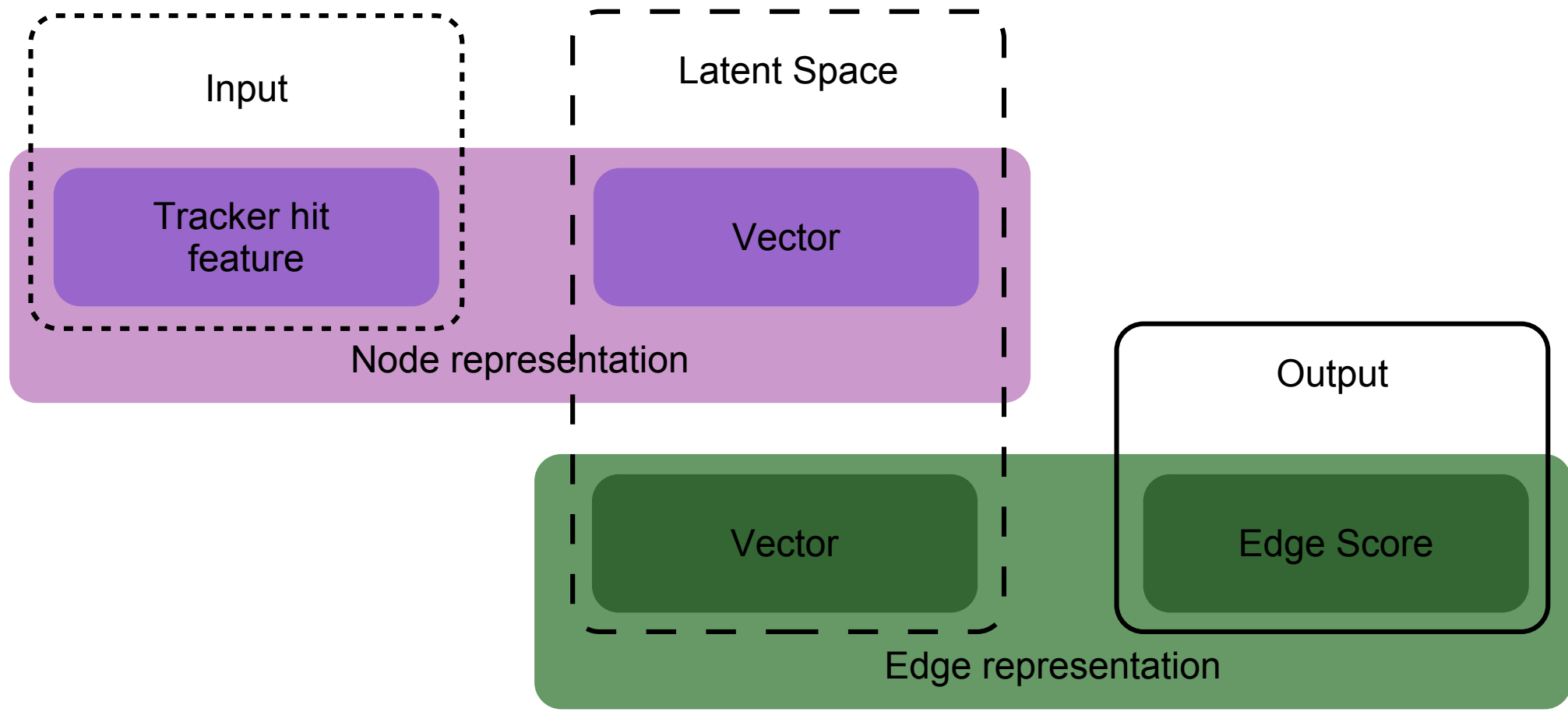
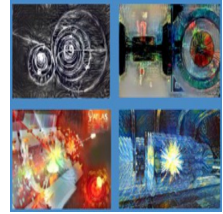


Grey edge: fake
Red edge : true





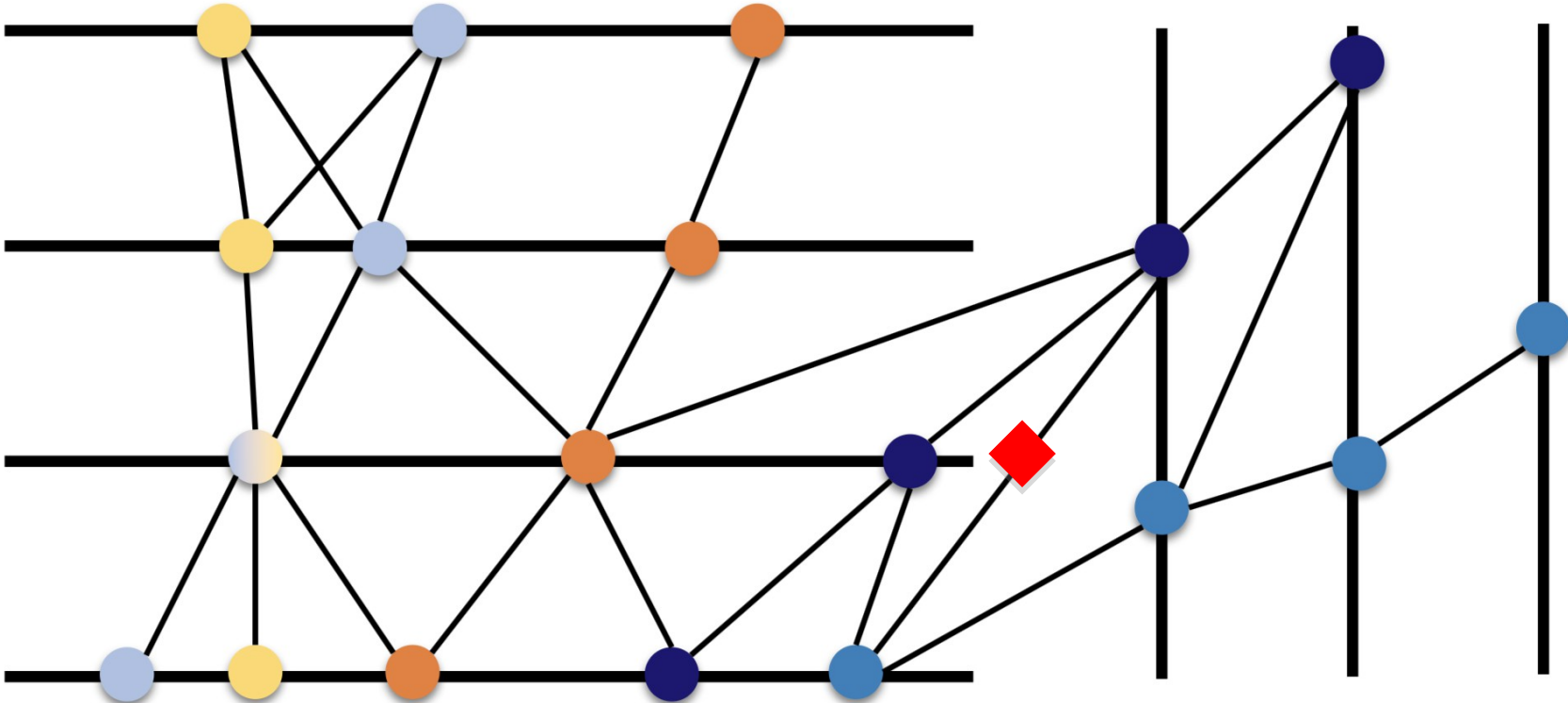
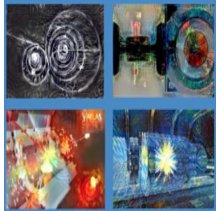
Node & Edge Representations



Edge representation is not the edge score.
Final edge score extracted from the latent edge representation.



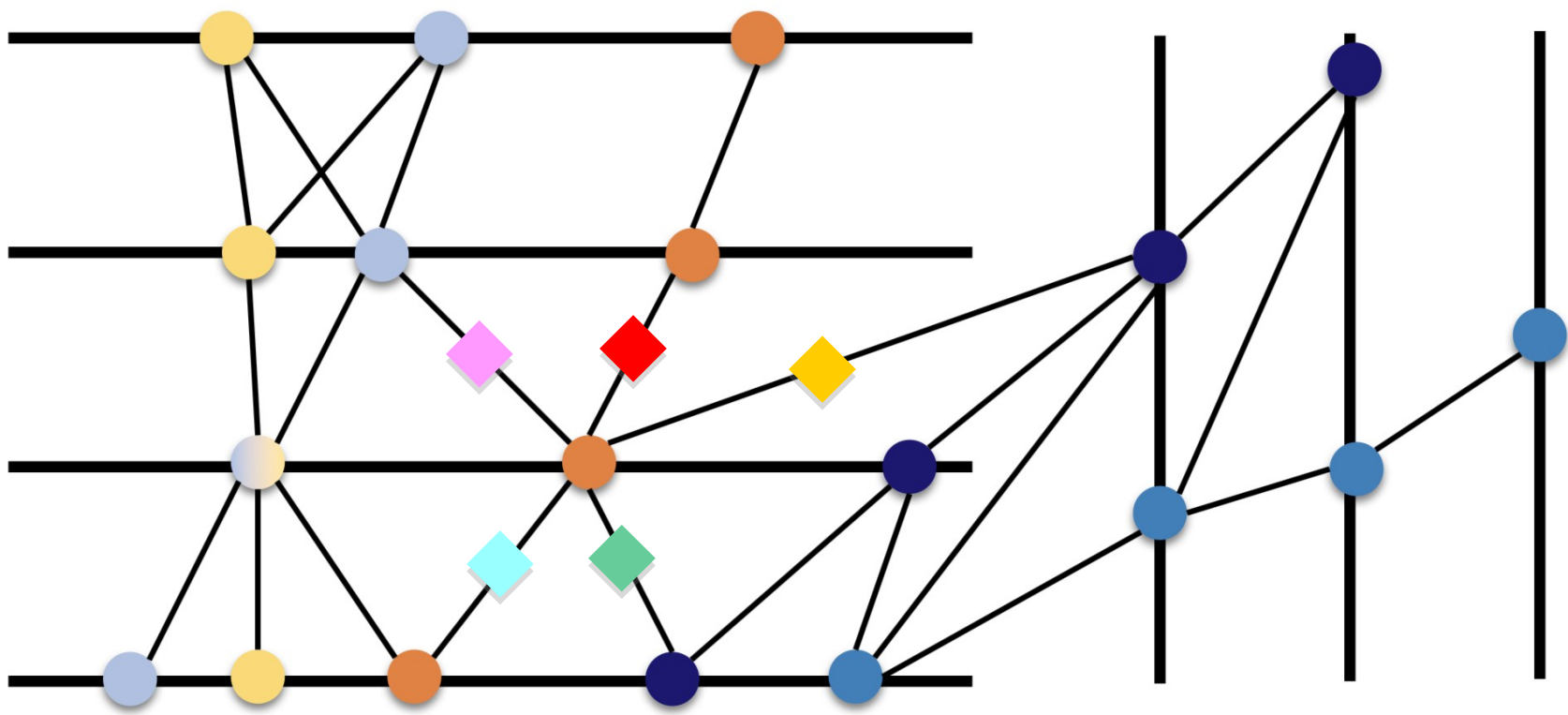
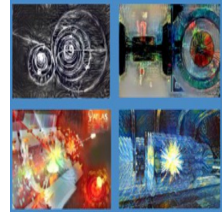
Edge Network



◆ ← EdgeNet(●, ●)



Node Network

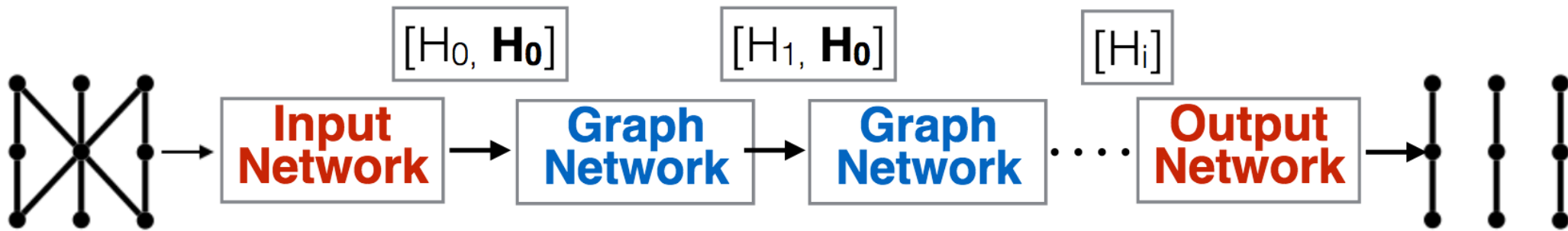
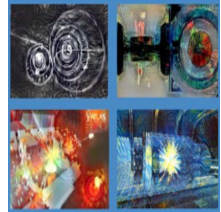


● ← NodeNet(●, ◆ + ◆ + ◆ + ◆ + ◆)

self connecting



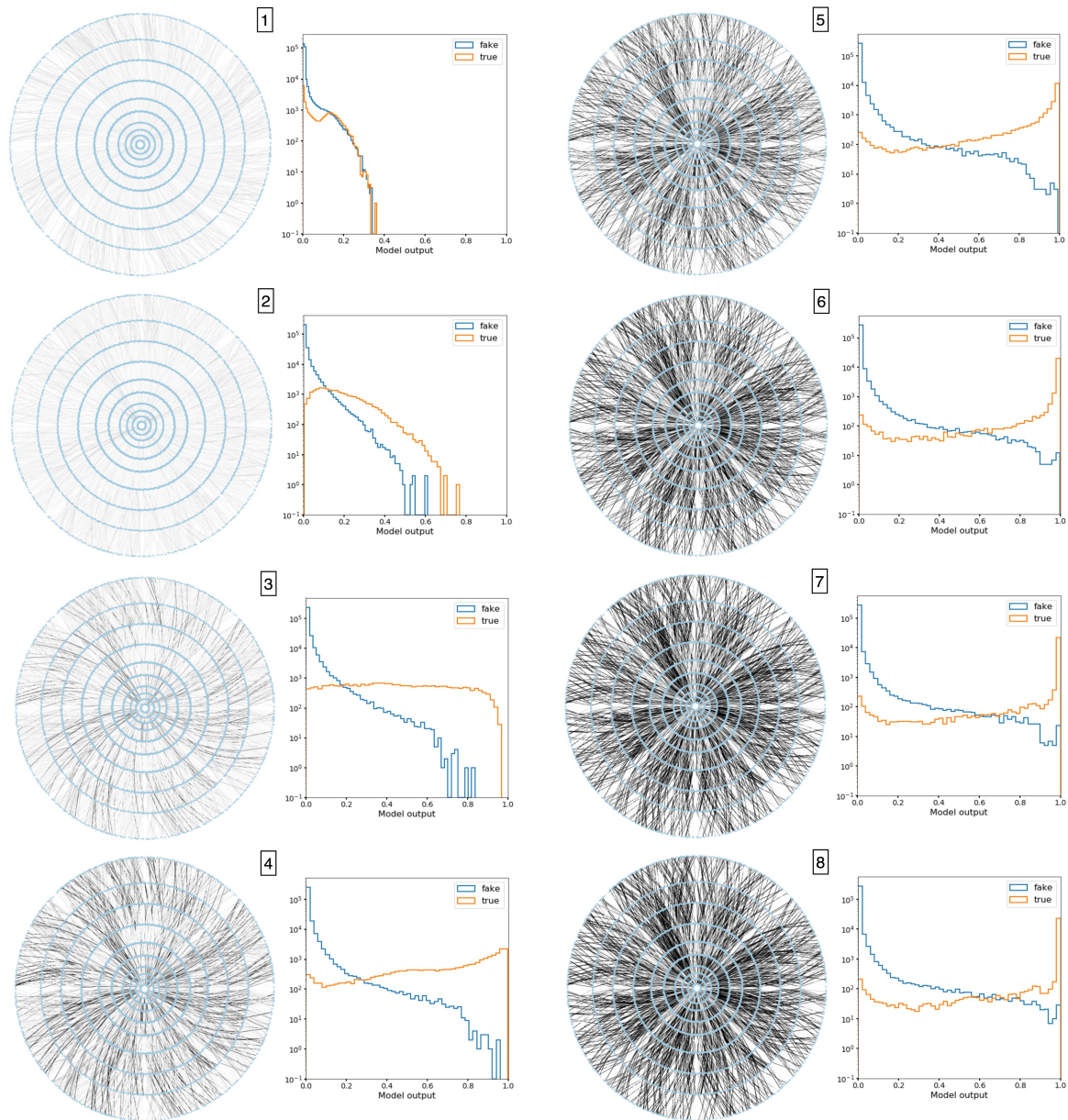
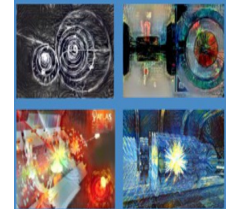
Message Passing Model



- Graph is sparsely connected from consecutive layers
 - Edge representation computed from features at the ends
 - Node representation computed from the sum over all connected edges
- Correlates hits information through multiple (8) iterations of (Graph Network)
- Uses https://github.com/deepmind/graph_nets TF library



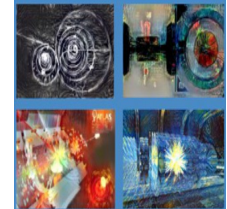
Information Flow



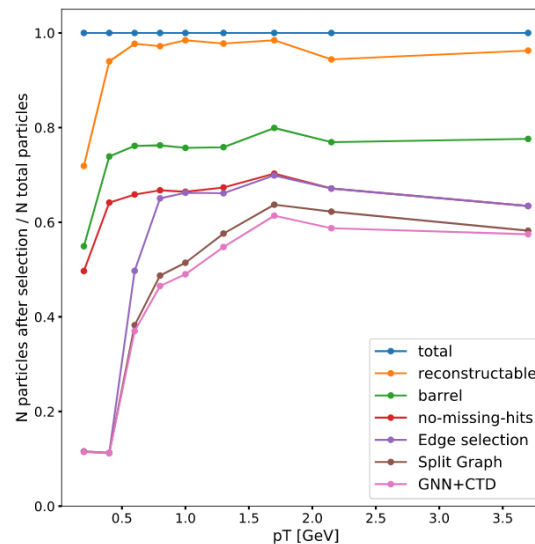
- Checking edge score after each step of graph network.
- Effective output of the model is in step 8.
- Full track hit assignment learned in last stages of the model.
- Tracklets learned in intermediate stages.



Performance



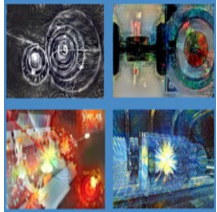
one-event	N-particles	ratio w.r.t Total	ratio w.r.t Reconstructable	relative ratio
Total	11170	100%		100%
Reconstructable	9635	86%	100%	86%
Barrel	7492	67%	78%	78%
No-missing hits	6600	59%	69%	88%
Edge selection	3114	28%	32%	47%
Split graph	2668	24%	28%	86%
GNN	2590	23%	27%	97%



- Tracks formed with a simple algorithms that traverse the hit graph over high-score edges.
- Promising performance, once passed acceptance cut for training purpose

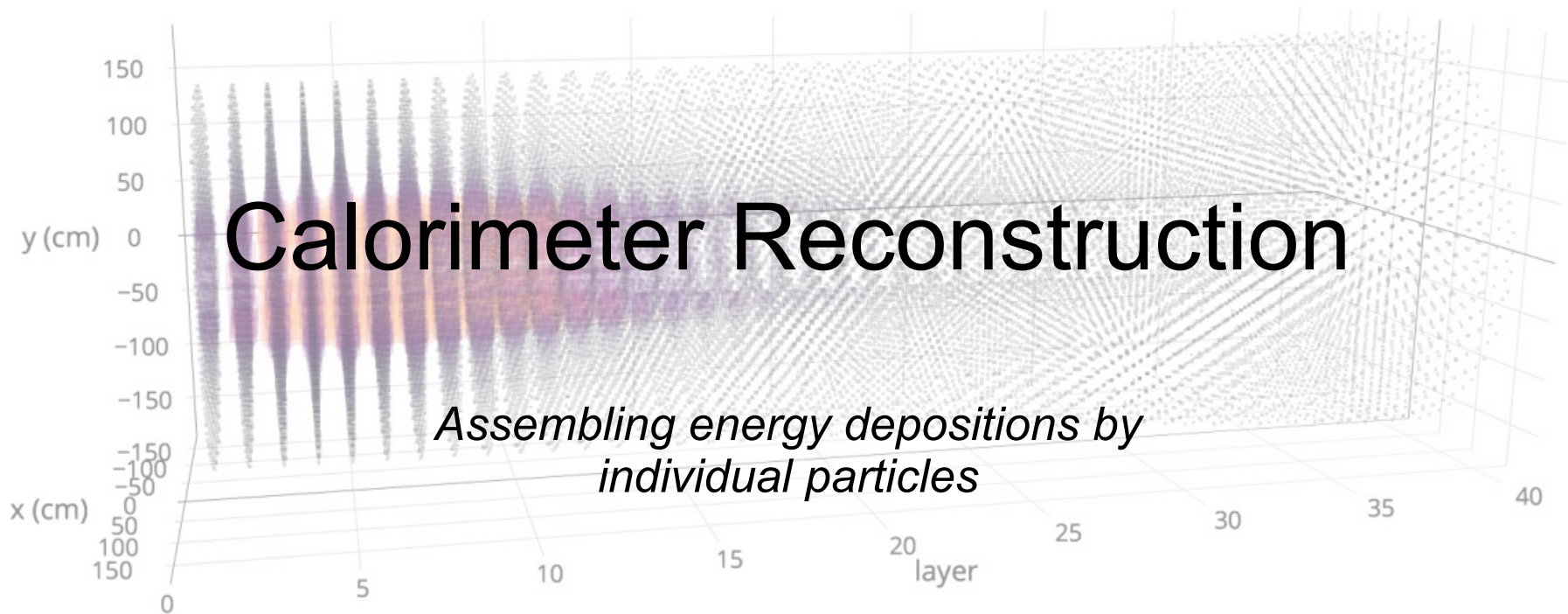
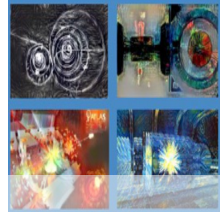


Remarks



- *The construction of the graph can limit the type of tracks that can be found.*
- *Over-connectivity produces highly unbalanced classes (many more fake edges than true).*
- *Computation graph can become cumbersome if the sparsity is not taken into account.*
- *What have we done with the Physics knowledge of particle dynamic ?*
- *Promising results but no solution yet ; stay tune for results from Exa.trkX.*

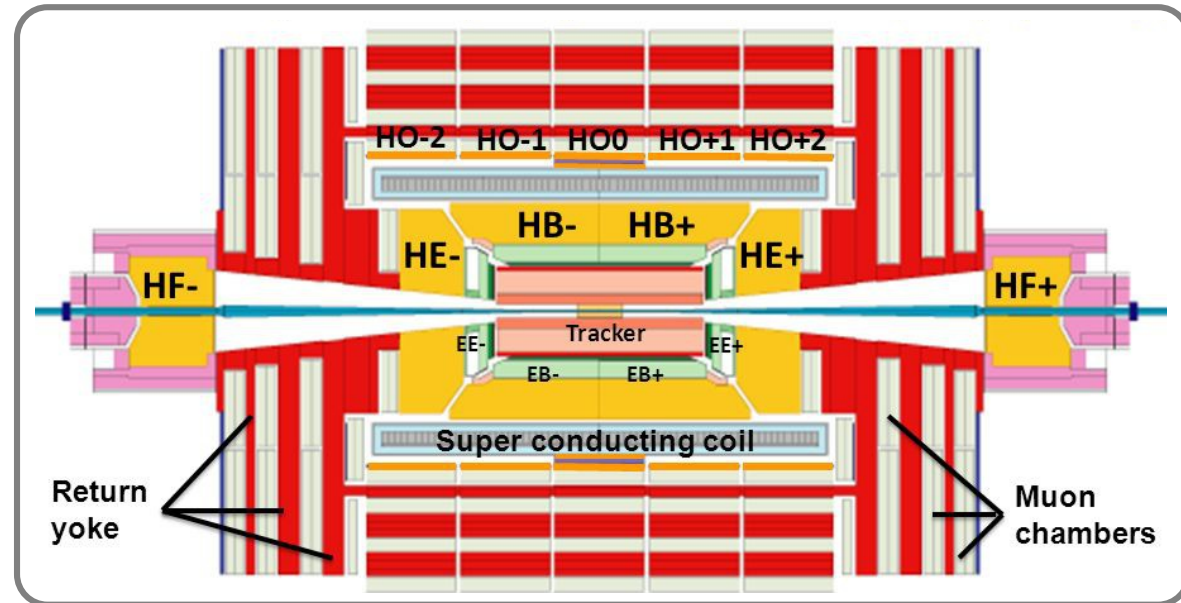
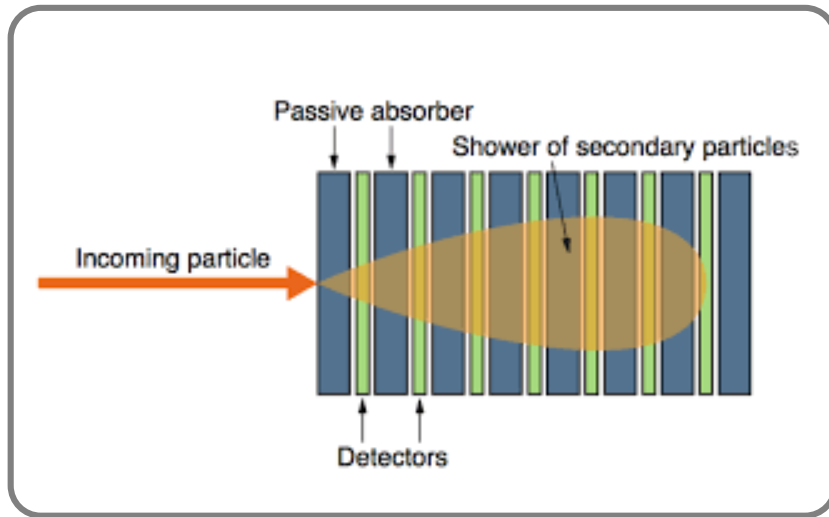
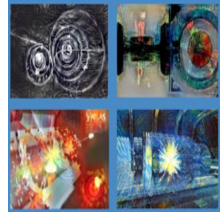
<https://github.com/HEPTrkX/heptrkx-gnn-tracking>
<https://github.com/xju2/heptrkx-gnn-tracking>



by Shah Rukh Qasim, Jan Kieseler, Yutaro Liyama, Maurizio Pierini
<https://arxiv.org/abs/1902.07987>



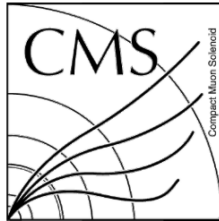
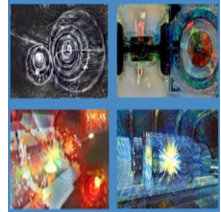
Calorimetry – 101



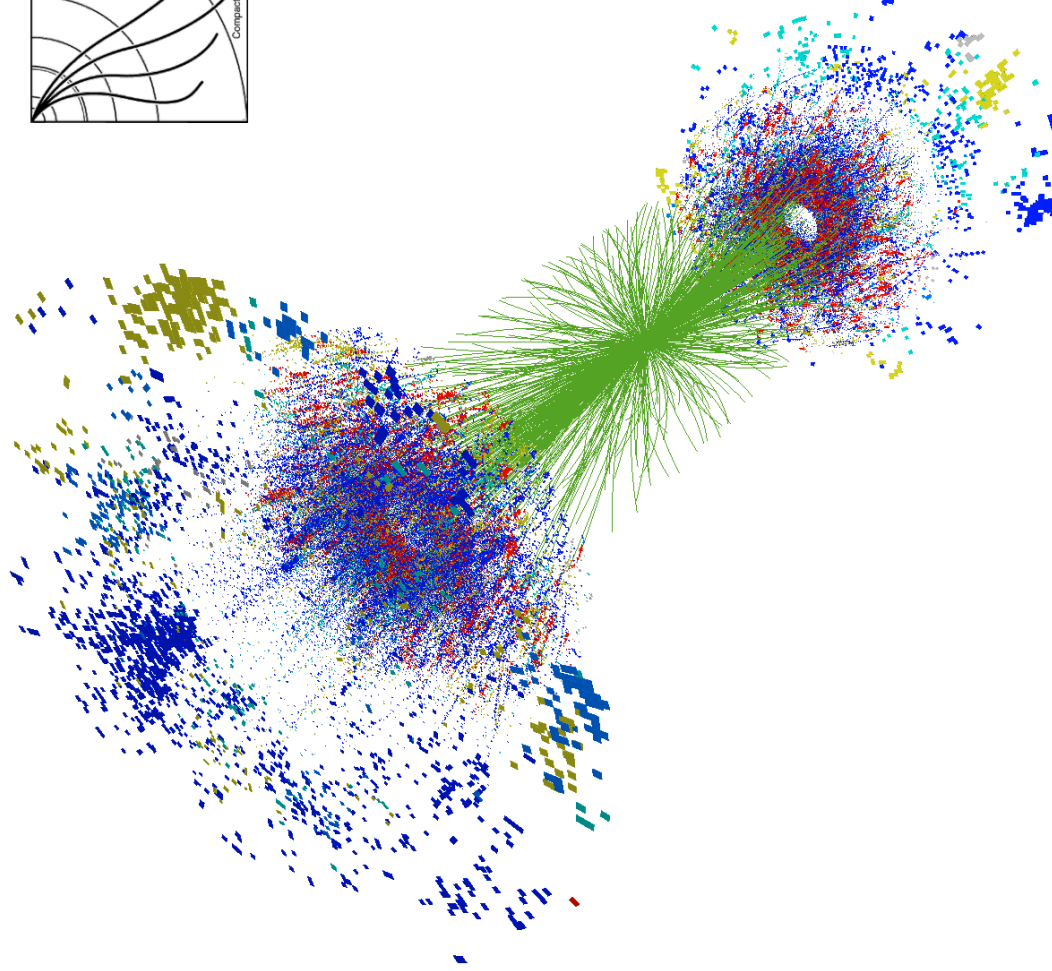
- Induced particle shower through material
- Destructive/stopping interaction with matter
- Sensitive cells measure properties of secondary particles
- Multiple technology available for sensitive material
- Assembled to hermetically cover most of all particles



Name of The Game



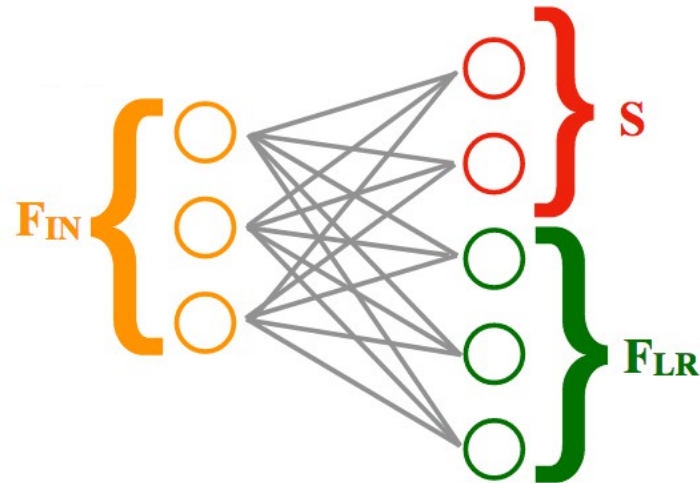
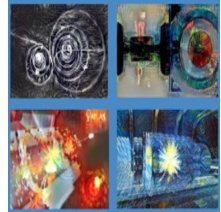
Ground truth color coded event display of the future high granularity calorimeter.



Associate each energy deposition from the secondary particles to their primary particle



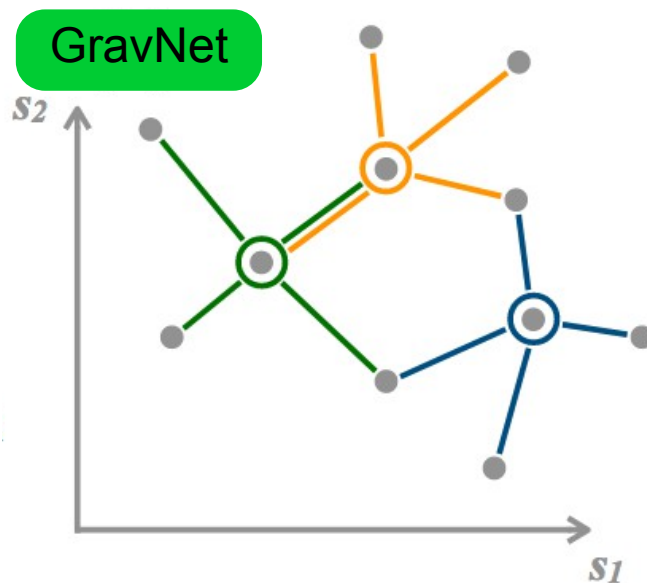
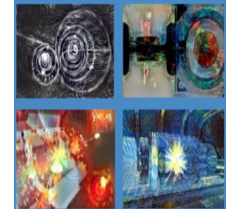
Feature Encoding



The calorimeter cell features are transformed into a **locations/distances** interpreted in some abstract space and an **internal representation**, using a fully connected network.



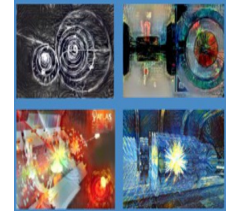
Abstract Space – GravNet



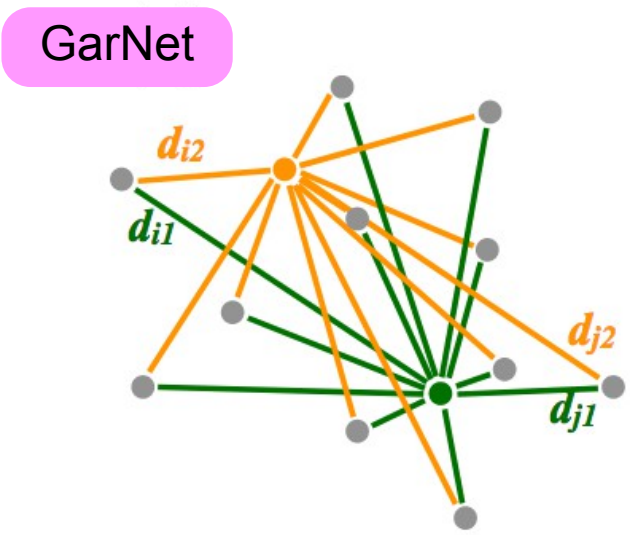
GravNet: graph defined from N nearest neighbours using **S** as location in abstract space.



Abstract Space – GarNet

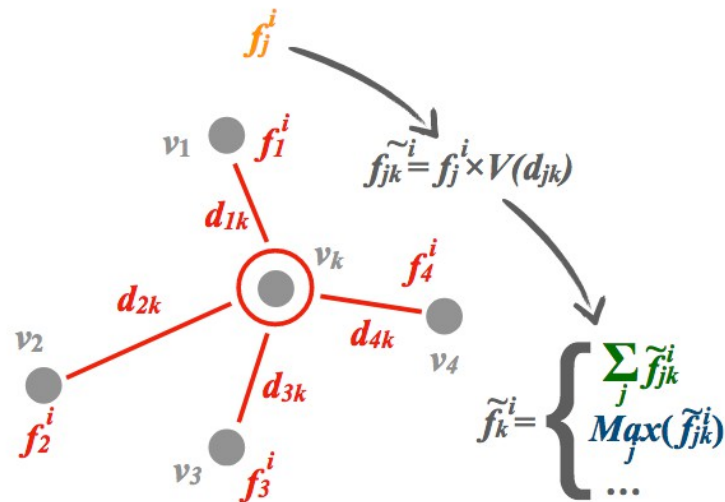
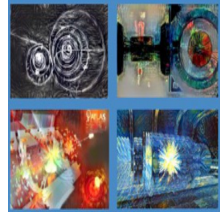


GarNet: graph defined with extra nodes using **S** as distance to attractors in abstract space.





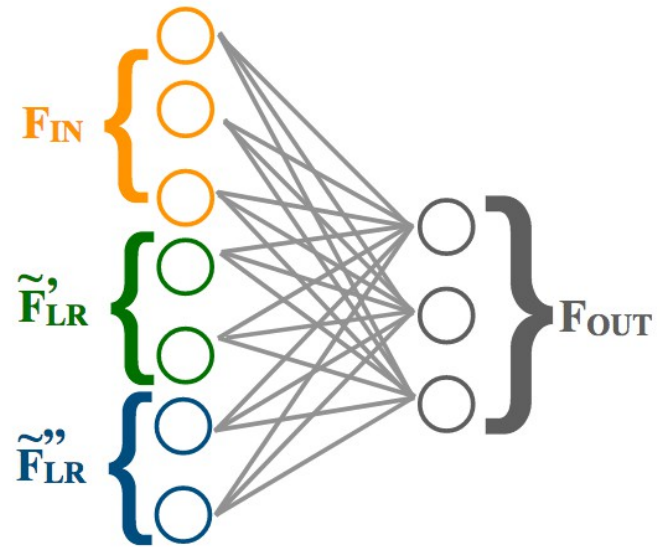
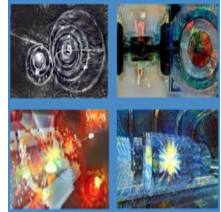
Internal Representation



The internal representation are weighted using a potential of the distances through each edge ($V(d_{jk})$), and new representation is calculated from **the mean** or **the max**.



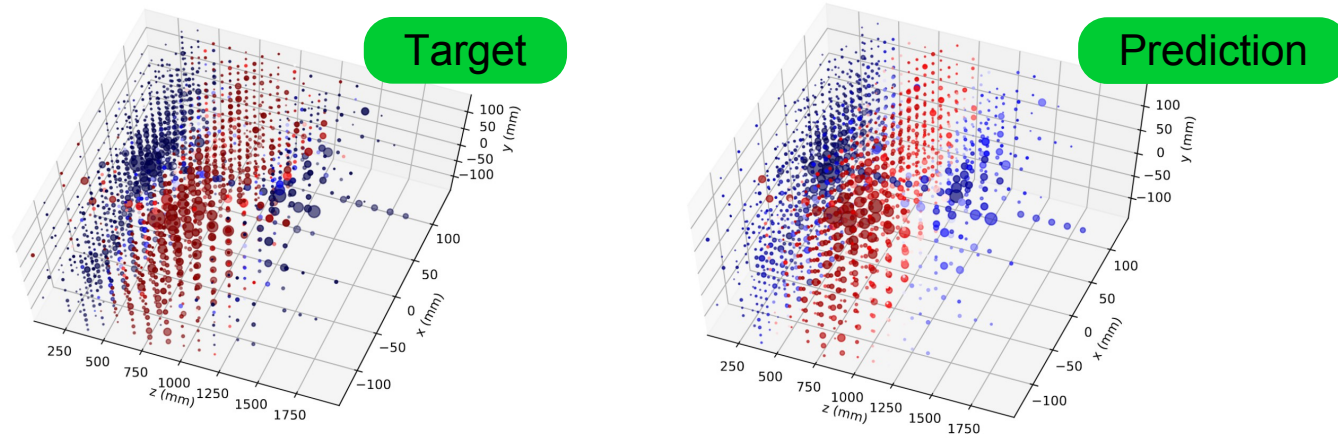
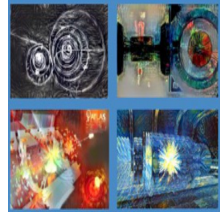
Feature Extraction



Predictions are extracted per calorimeter cell, from **initial features** and gathered **internal representations**, using a fully connected network.



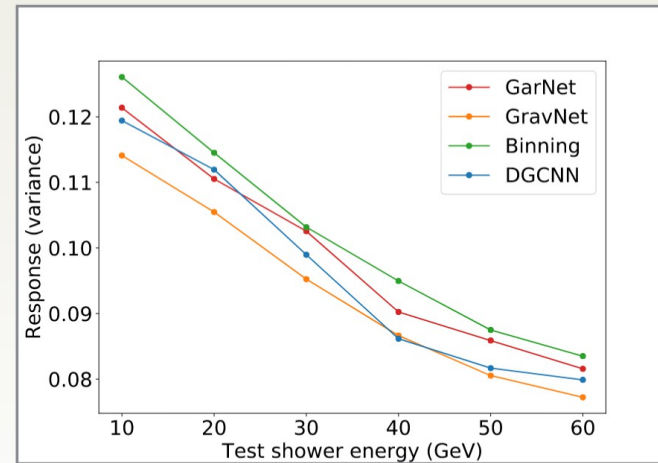
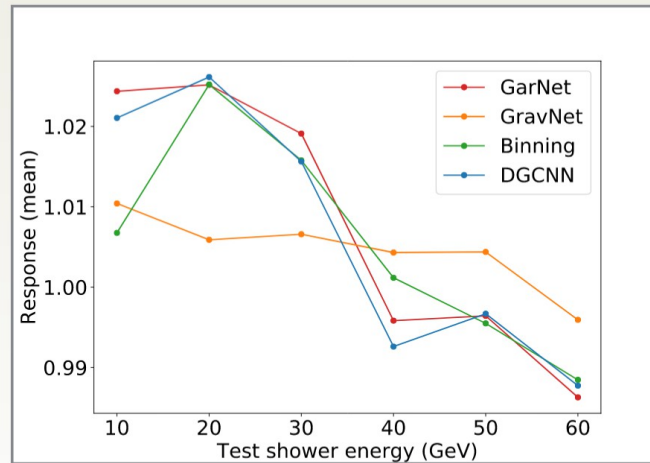
Performance



Overlapping showers from two charged pions.

- Focus on the overlap region, only (20-80 % overlap)
- Define energy response

$$R_k = \frac{\sum_i E_i p_{ik}}{\sum_i E_i t_{ik}}$$

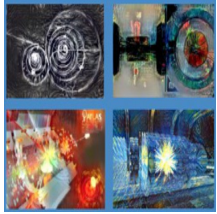


- The graph network based approaches outperform the CNN approach
- The GravNet model outperforms all approaches

Slide J.Kieseler

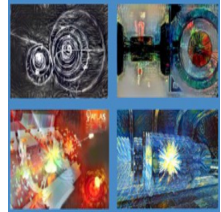


Remarks



- *Graph representation seems to be appropriate for calorimeter with complex geometry.*
- *Graph connections may be dynamic (k- nn) as long as the computation graph remains differentiable.*
- *Node representation interpreted in abstract space to match the needs.*
- *People are also working on “Edge classification” models*
- *What have we done with the Physics knowledge of particle shower dynamic ?*

<https://github.com/jkieseles/calographnn>



CMS
E
CMS Experiment at LHC, CERN
Data recorded: Mon May 28 01:16:20 2012 CEST
Run/Event: 195099 / 35438125
Lumi section: 65
Orbit/Crossing: 16932111 / 2295

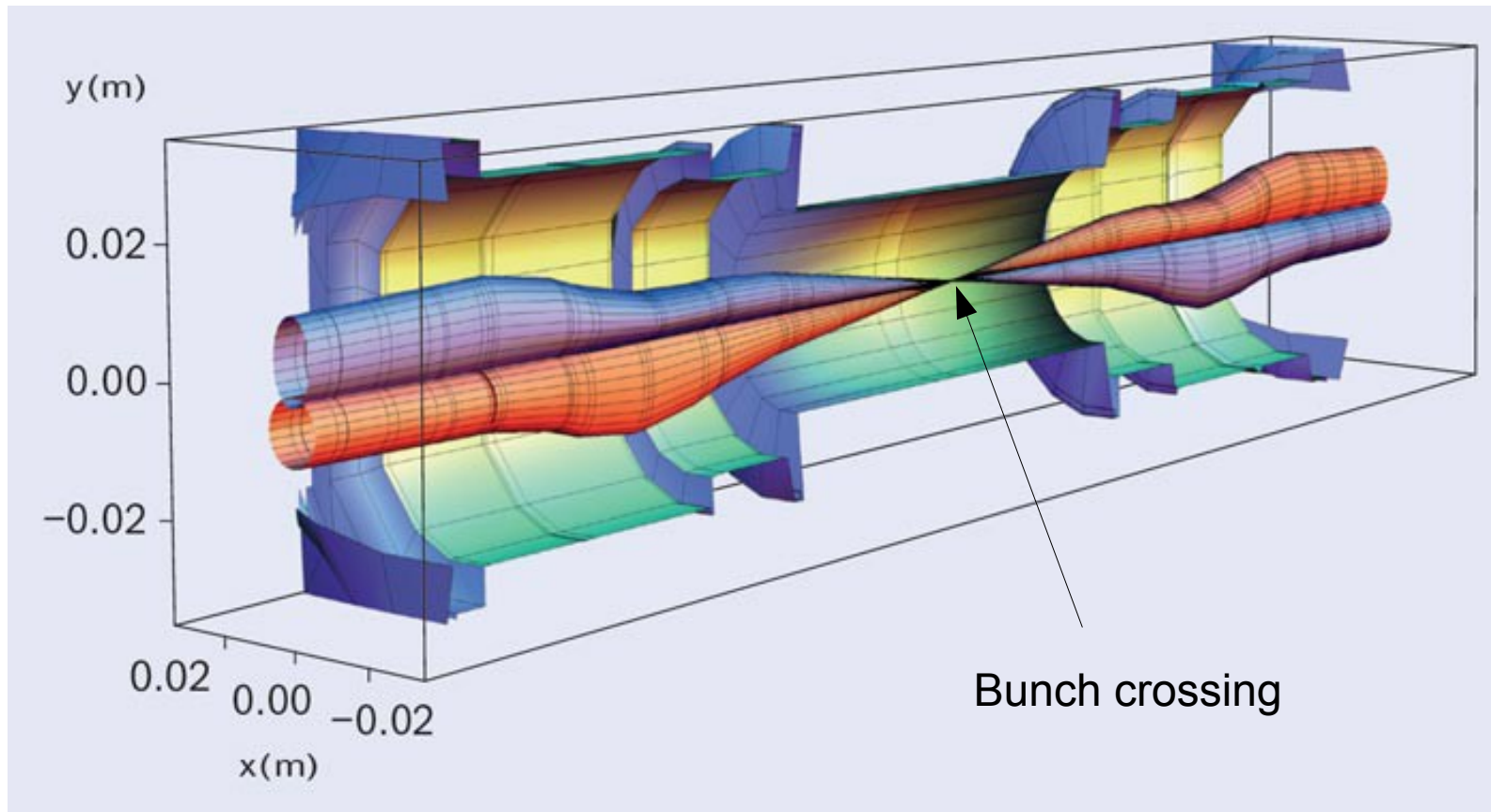
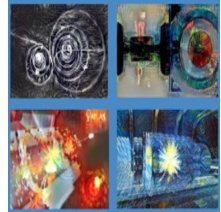
Pile-up Mitigation

Identifying/rejecting particles coming from parasitic proton-proton collisions.

With Jesus Arjona Martinez, Olmo Cerri, Maurizio Pierini, Maria Spiropulu
<https://arxiv.org/abs/1810.07988>



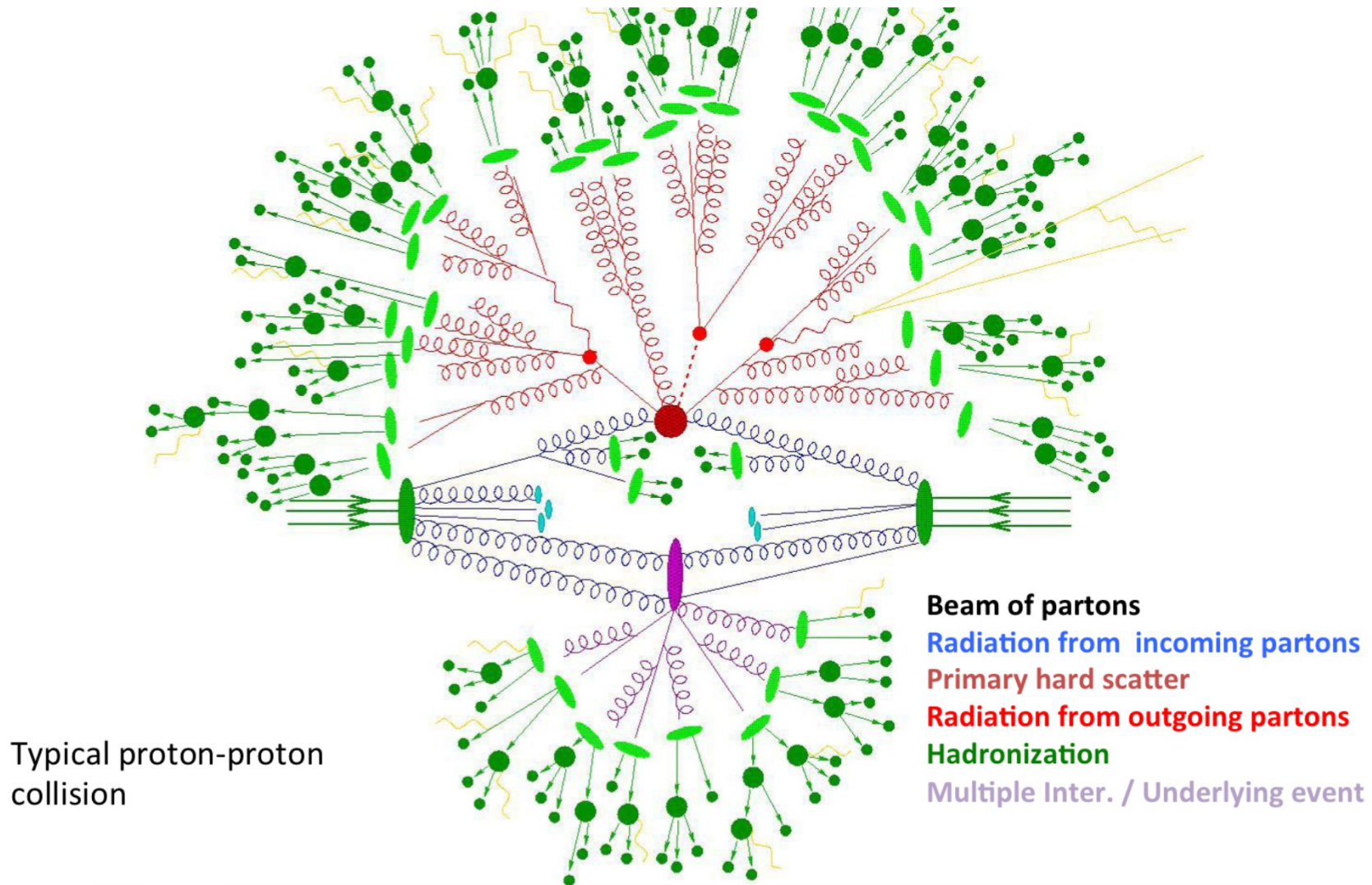
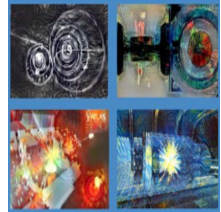
Beam Crossing – 101



- Rate of interactions can be driven up
 - Bunch crossing every 25 ns (40MHz)
 - Density of protons per bunch
- Concurrent interactions is the simplest route for increased rate of interactions



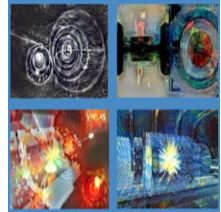
Pileup – 101



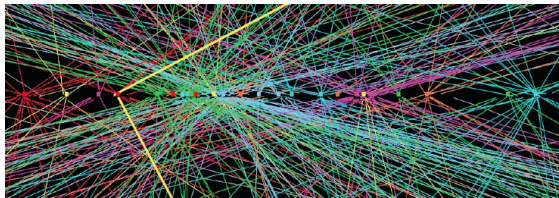
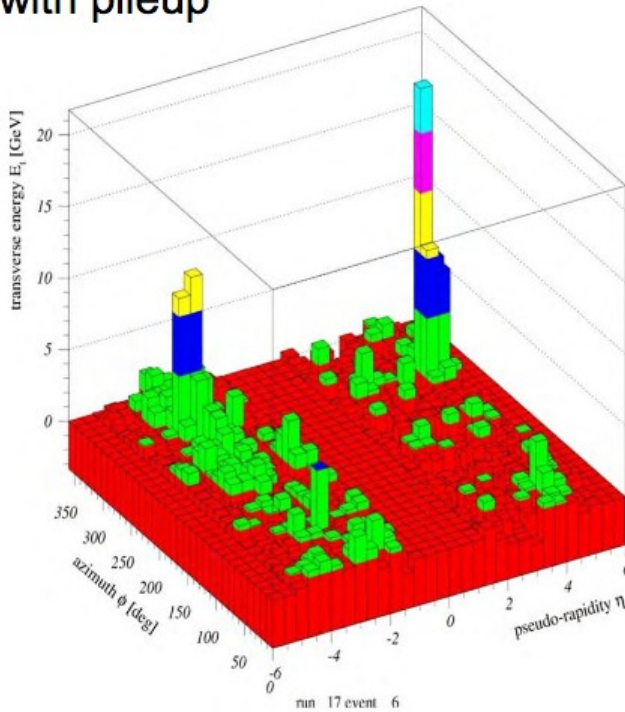
Bunch crossing every 25 ns / 40MHz \equiv one event.
Multiple collisions per bunch (pile-up) for increased probability.
200 averaged pileup in the horizon 2025.



Name of the Game



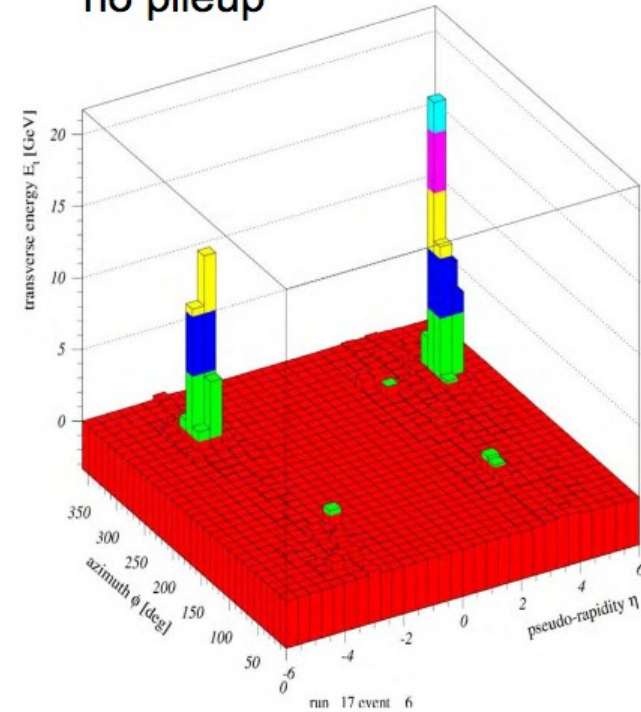
with pileup



Neutral hadron?

Charged hadron removed using track/vertex

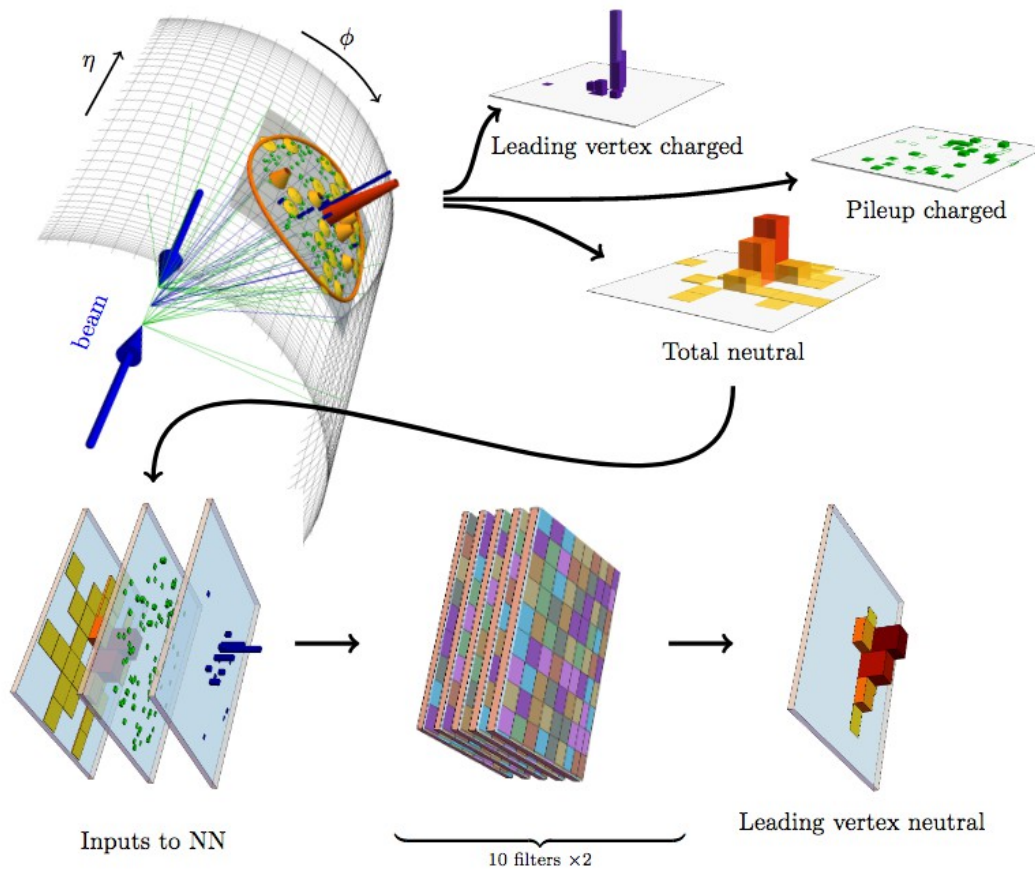
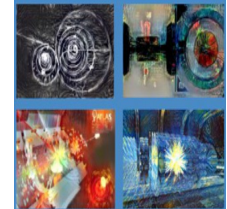
no pileup



Identify & remove the contribution of **neutral particles** coming from parasitic interactions.



Other ML Solutions

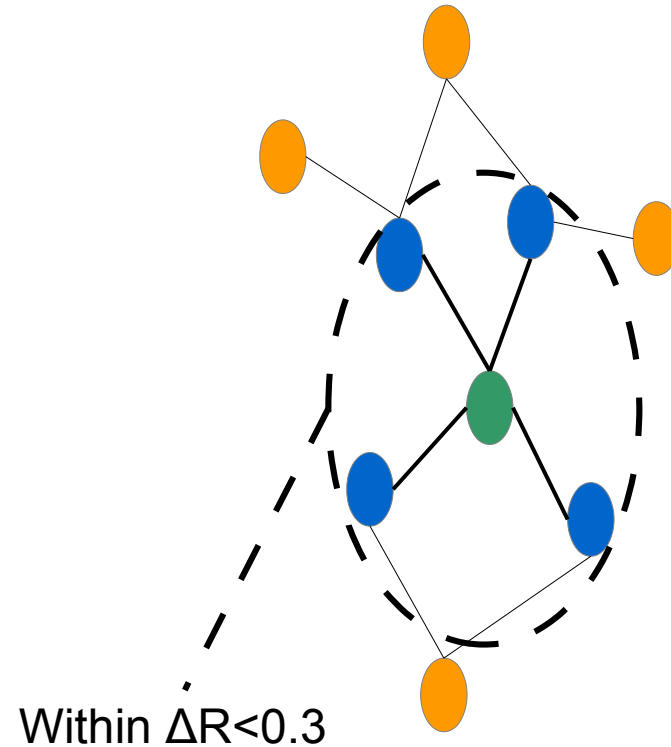
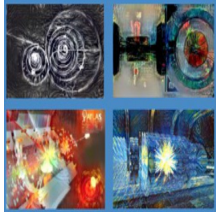


PUMML

<https://arxiv.org/abs/1707.08600>



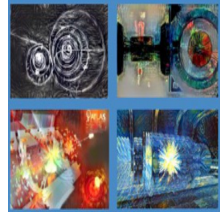
Graph Construction



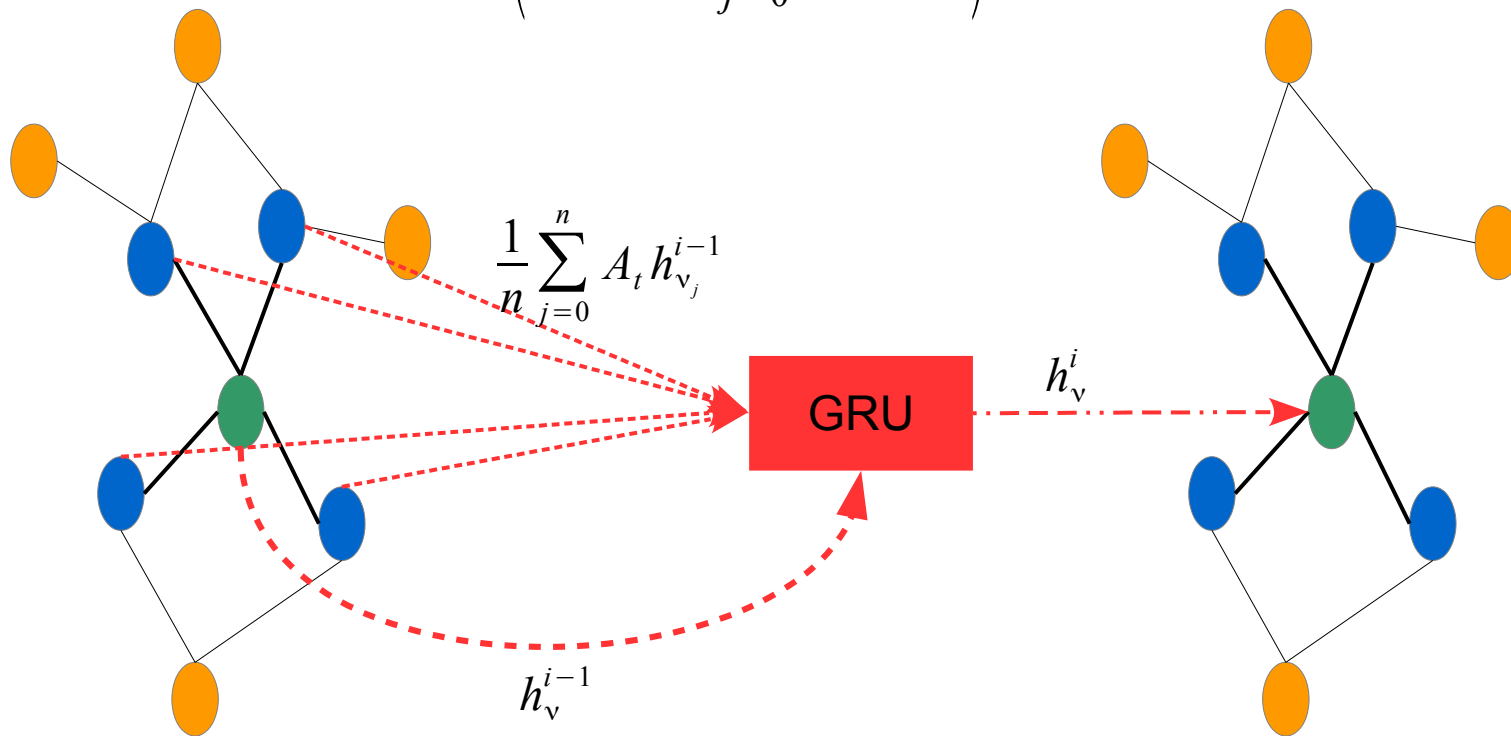
Graph constructed with **one particle** per node.
Edges of graph connecting **geometrically close particles**.
per-particle and global features assigned to nodes.



Updating Rule



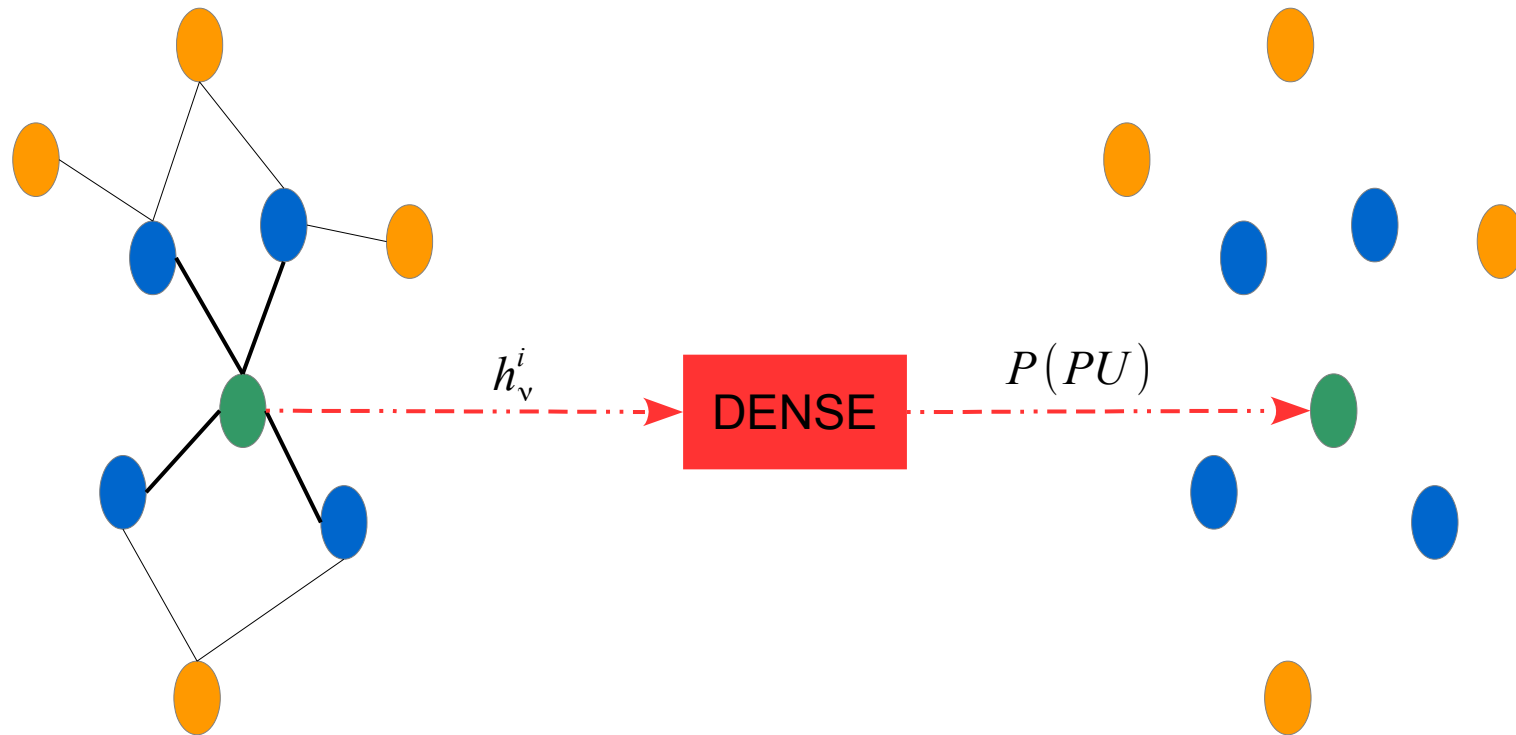
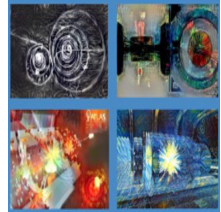
$$GRU \left(h_v^{i-1}, \frac{1}{n} \sum_{j=0}^n A_t h_{v_j}^{i-1} \right) \rightarrow h_v^i$$



Hidden/internal representation of the each node/particle updated with gated recurrent unit (GRU) models



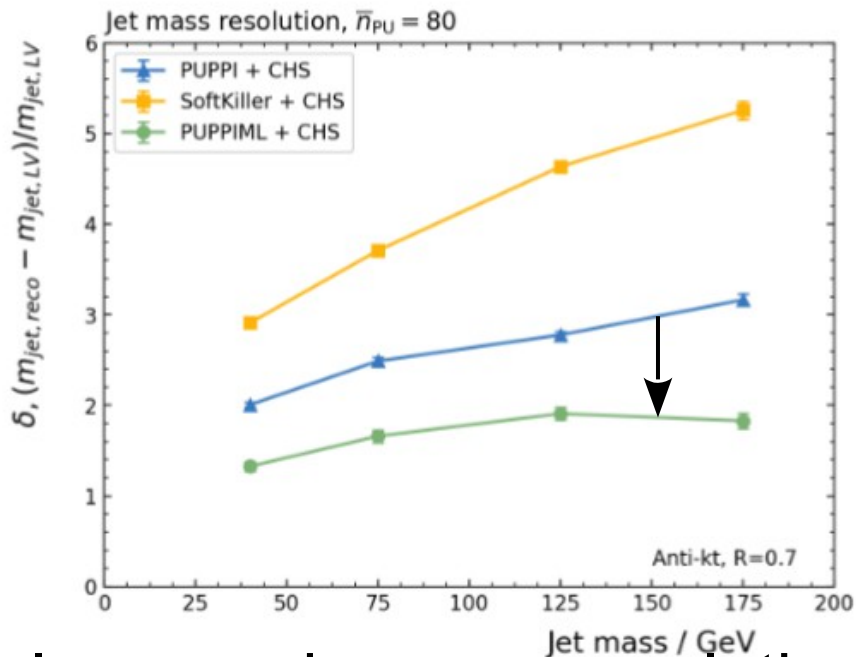
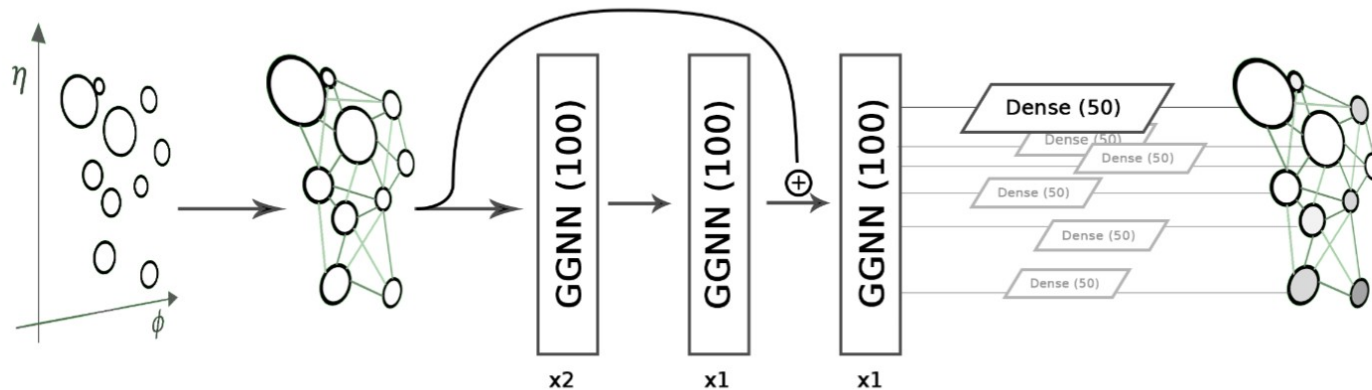
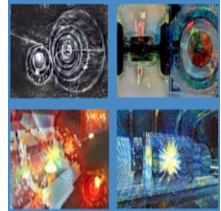
Pile-up Classifier



Binary classification extracted from the hidden/internal representation of the each node

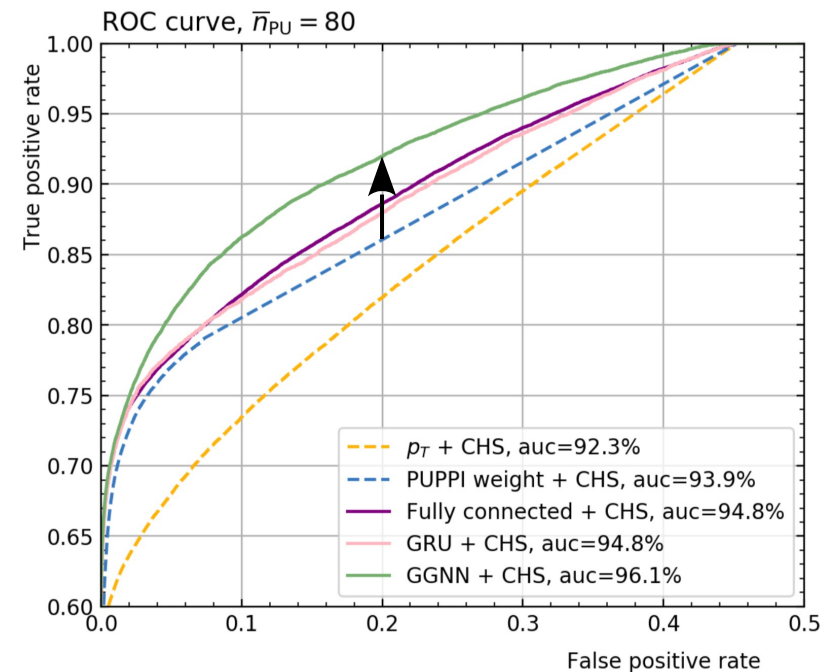


Model Performance



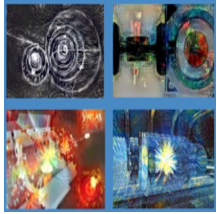
Improved energy resolution over state of the art pile-up removal methods.

Better rejection of underlying parasitic collisions.



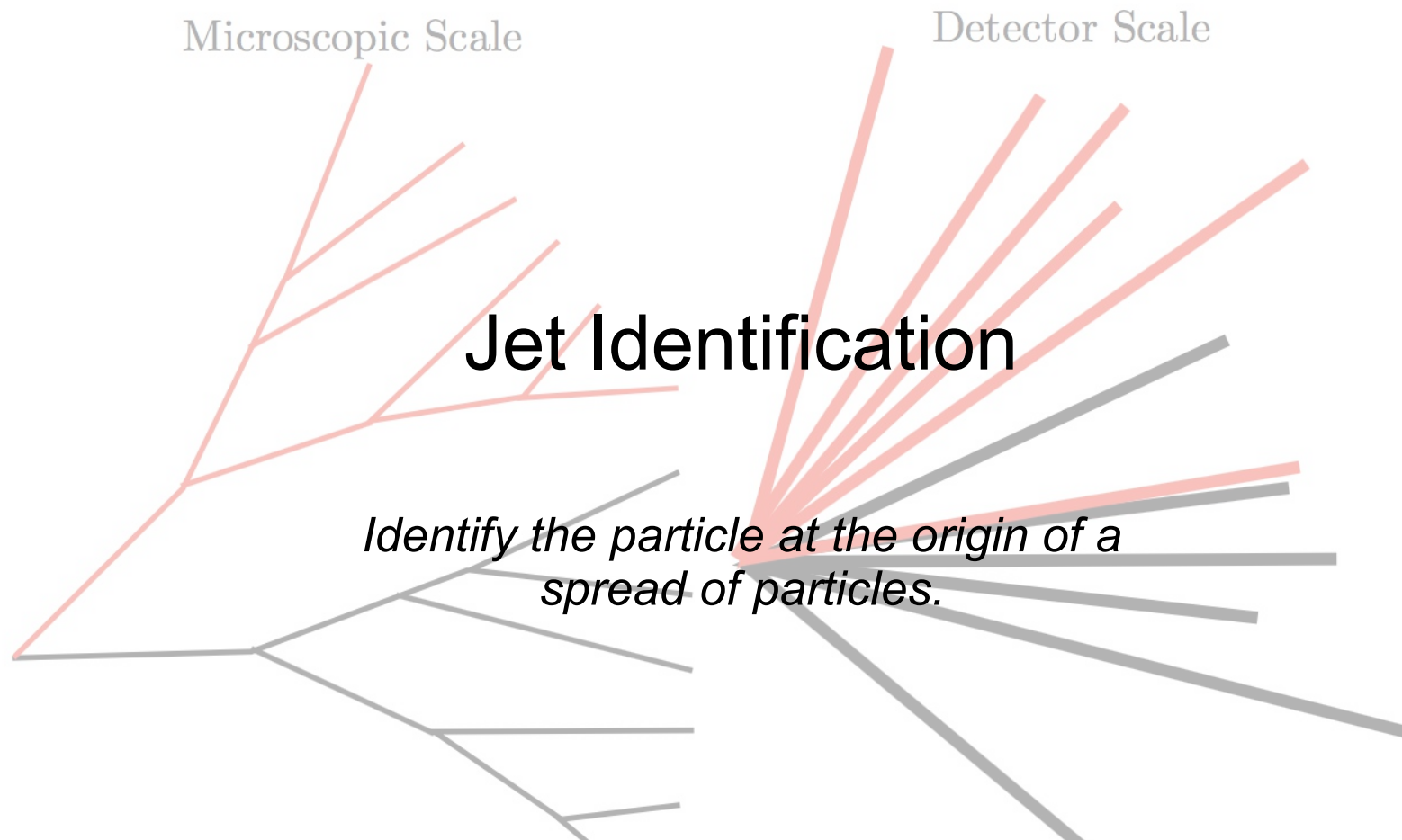
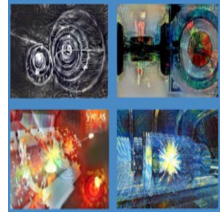


Remarks



- *Handling of global features in addition to per-particle features. In early stage, could be injected at later stage.*
- *Possibility to setup as a one-sided denoising algorithm.*
- *Can this be trained on data?*

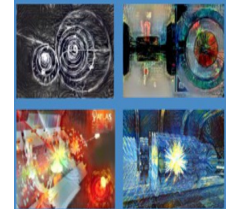
https://github.com/jarjonam/PUPPI_ML
will turn public “soon”



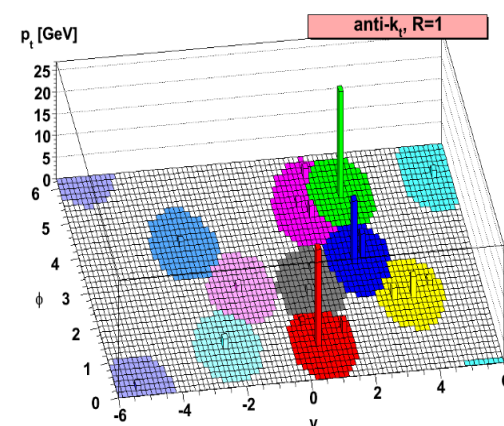
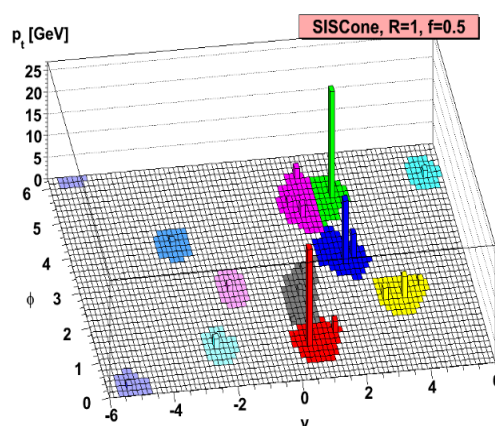
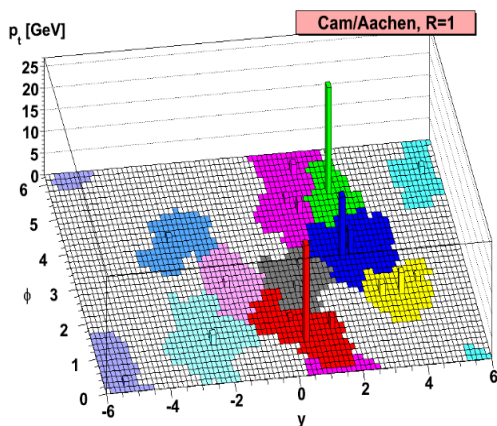
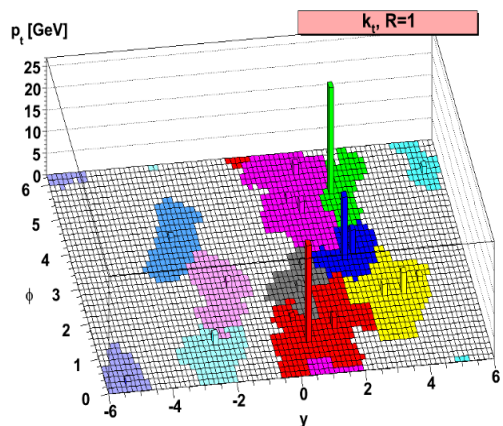
With Eric A. Moreno, A. Periwal, Olmo Cerri, Javier M. Duarte, Harvey B. Newman, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulu



Jet Reconstruction – 101



Clustering of particles/objects based on a distance metric.
Physics content/properties depends on the metric.



Three members of a family of algorithms:

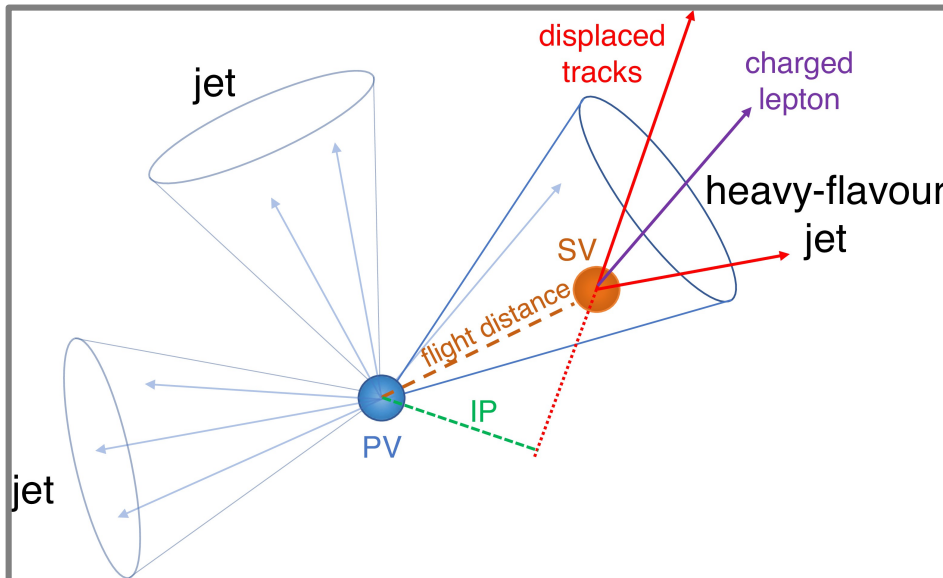
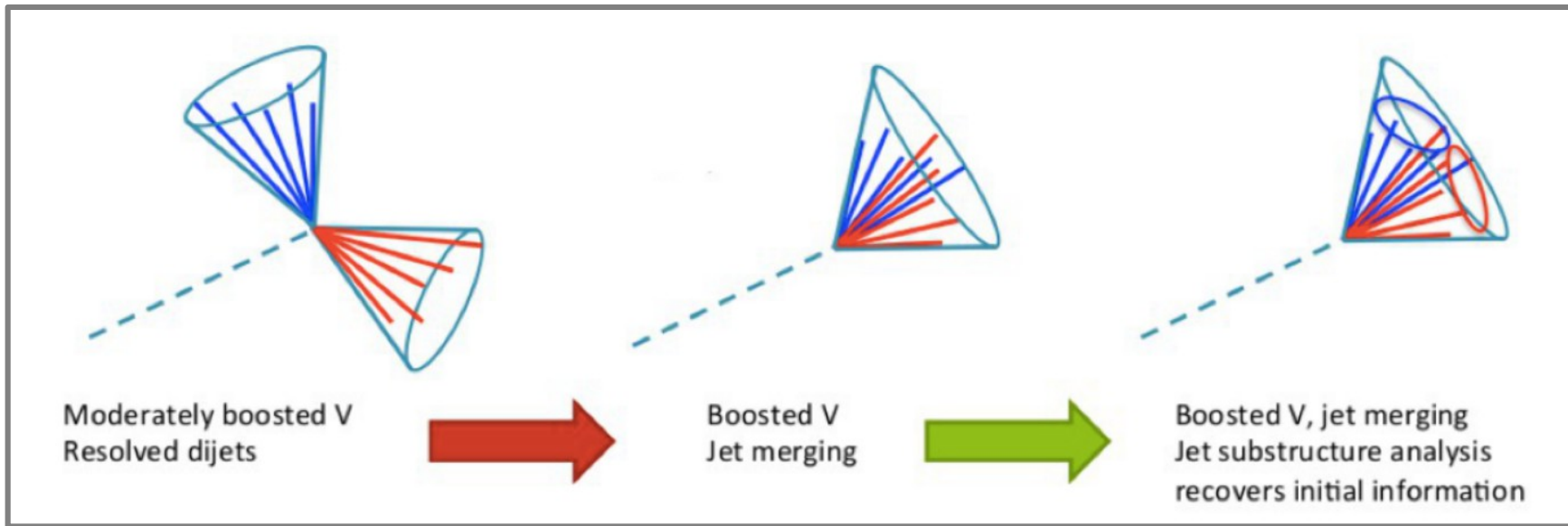
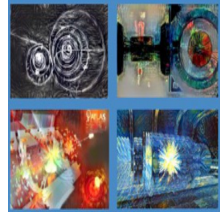
$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta y^2 + \Delta \phi^2}{R^2} \quad \begin{cases} p=1 & k_t \\ p=0 & \text{Cam/Aachen} \\ p=-1 & \text{anti-}k_t \end{cases}$$

Fastjet algorithm

<https://arxiv.org/abs/1111.6097>



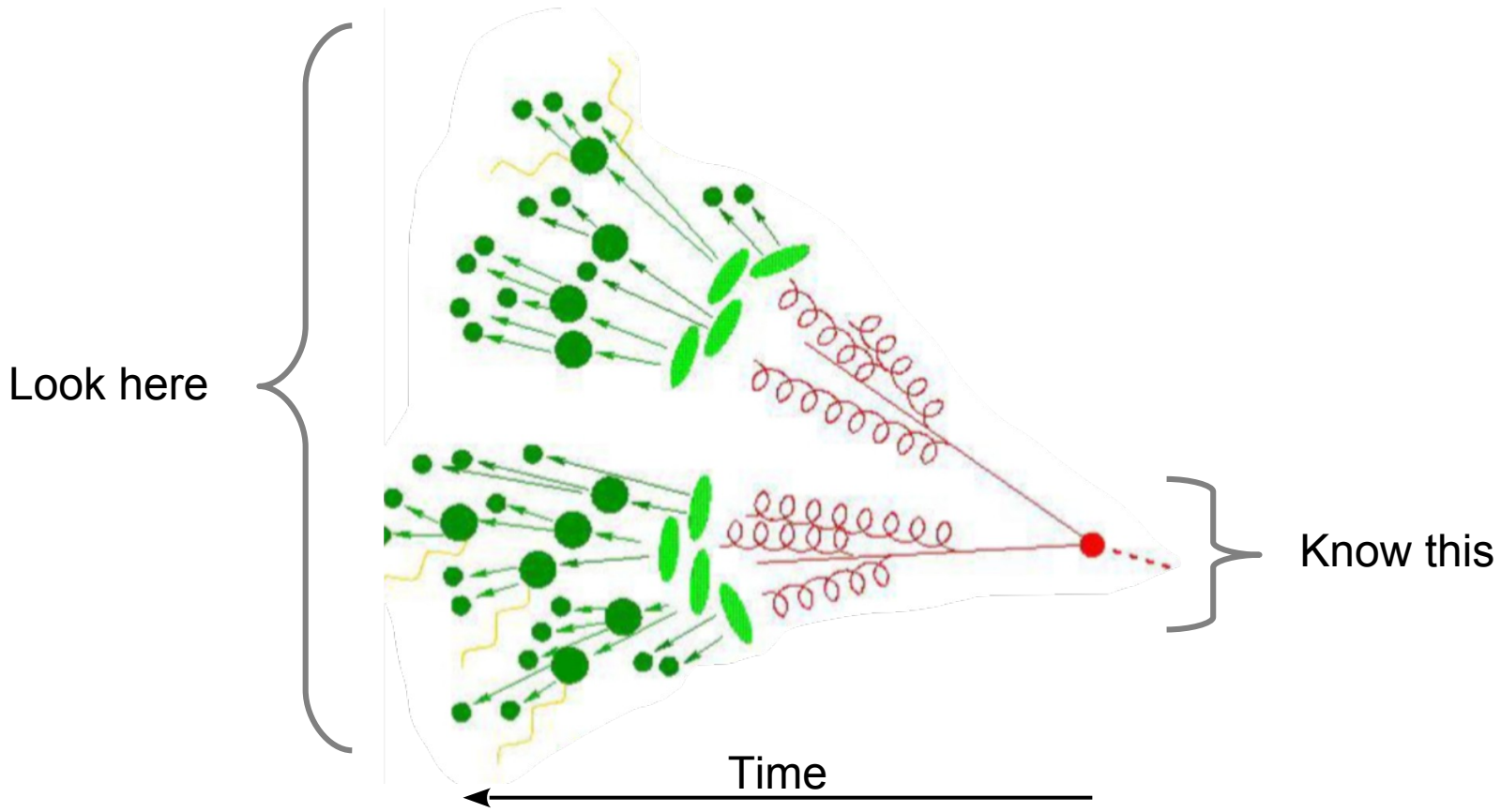
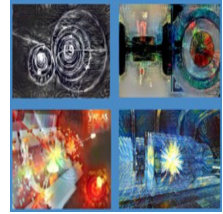
Jet Tagging – 101



- Secondary vertex is used to probe late decay of constituents
- Jet substructure derived from particle constituents used in boosted topologies



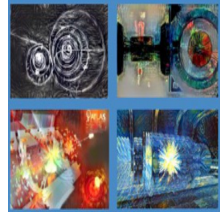
Name of the Game



Tell the characteristics of the particle at the origin of a bundle of secondary particles, looking at constituents and characteristics.

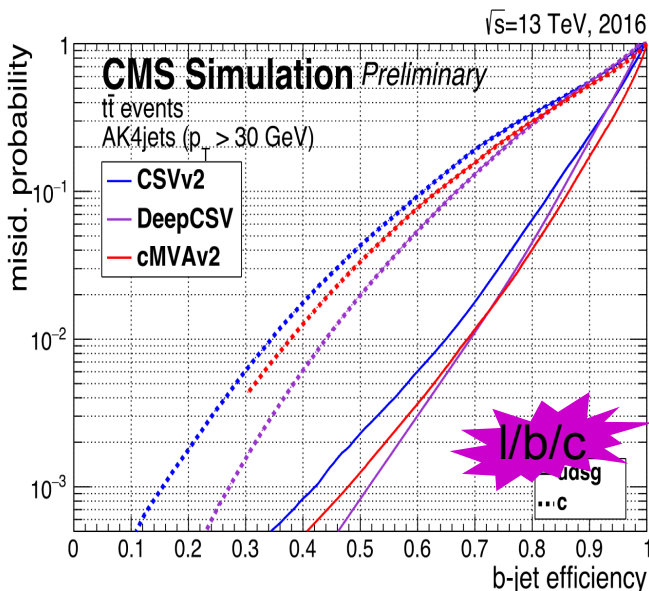
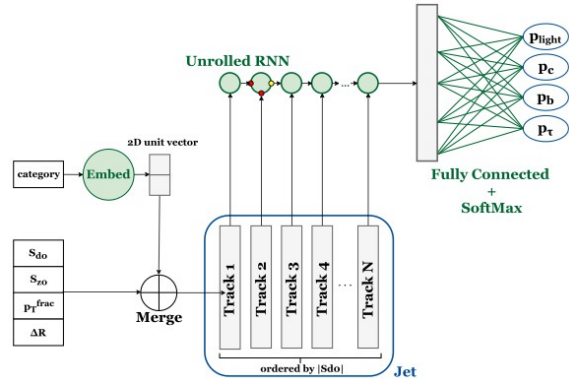


Other ML Solutions

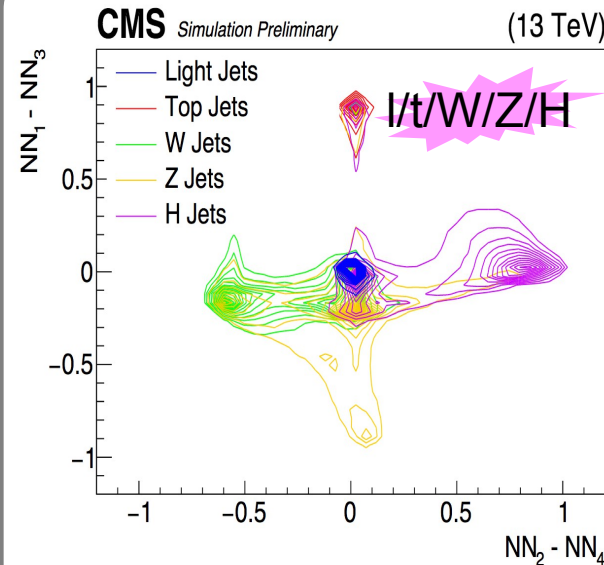


B-Jet with RNN

<http://cds.cern.ch/record/2255226>

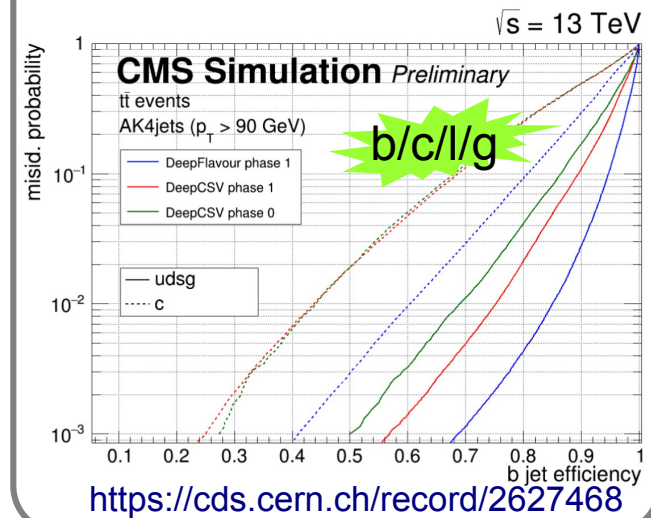


<http://cds.cern.ch/record/2255736>
 Combining jet-level, vertex-level and track-level features in FC NN.



<https://cds.cern.ch/record/2275226/>
 Jet-level features computed in hypothesized rest frame in dense multi-class neural net.

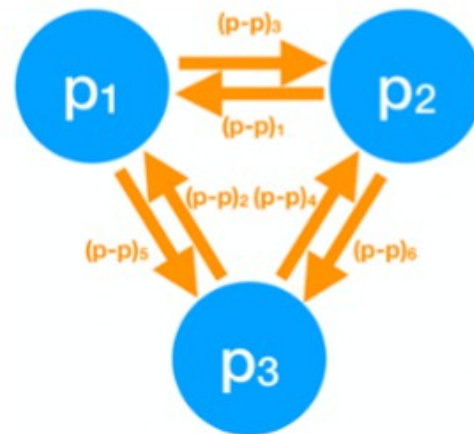
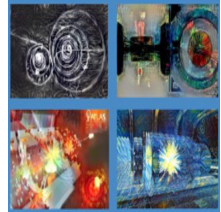
Jet-level, vertex-level and particle-level features in Conv1D/LSTM/FC multi-class neural net.



For more, see
<https://arxiv.org/abs/1709.04464>
<https://indico.cern.ch/event/745718/>



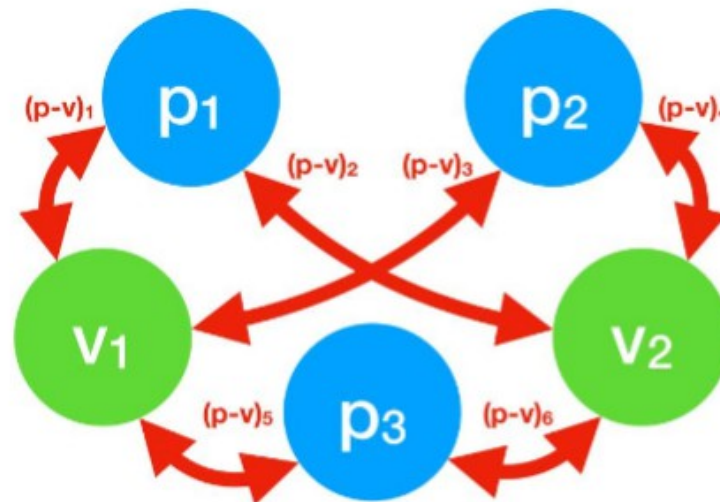
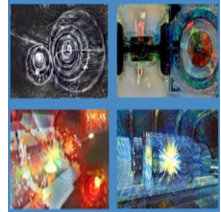
Particle Interaction



Learning interaction between particles with an all-to-all connected graph.



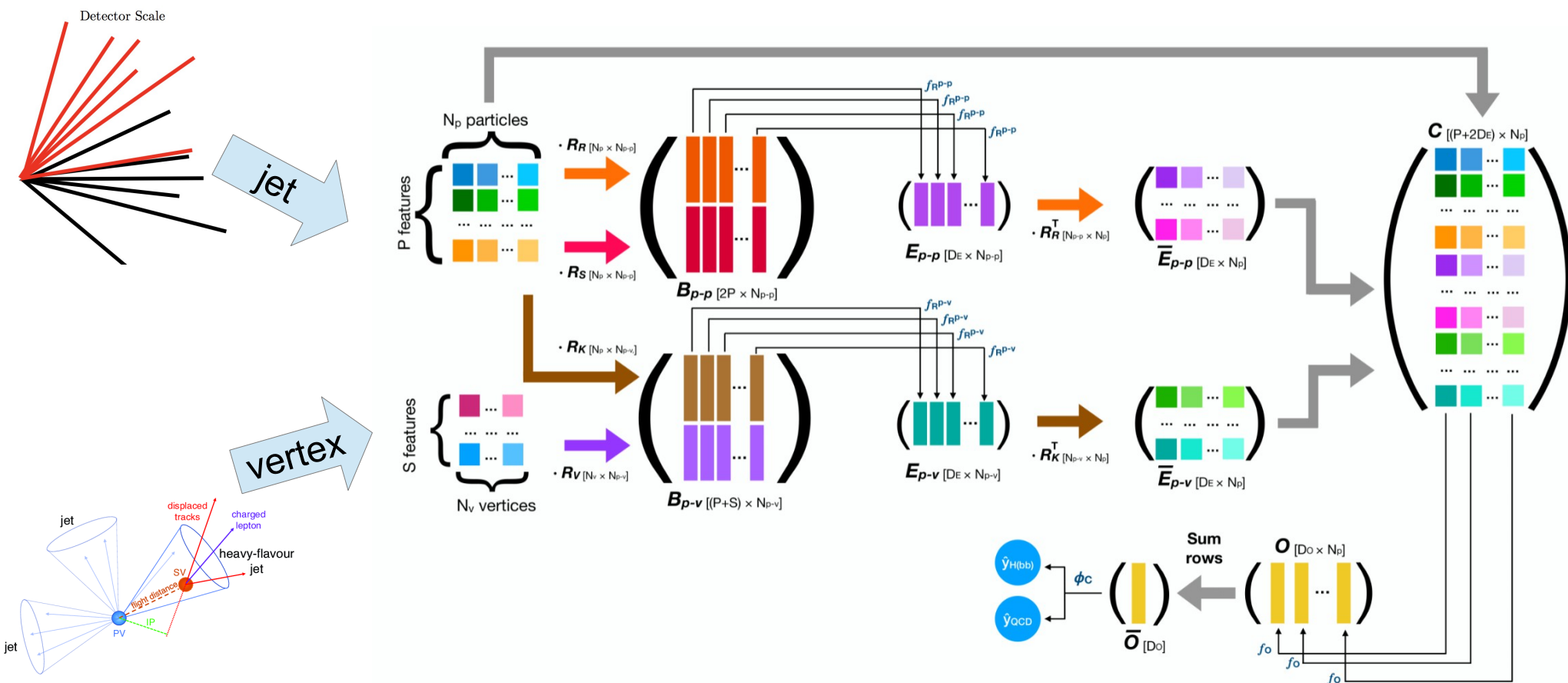
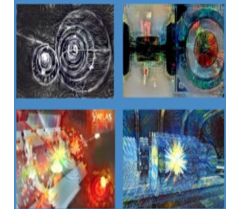
Particle/Vertex Interaction



Learning interaction between charged particles and vertex with all-to-all connected graph.



Jet-id with Graph Network

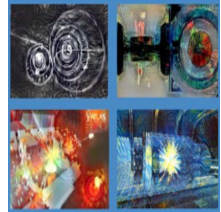


All particles of a jet, and vertex added on an all-to-all message passing graph network.

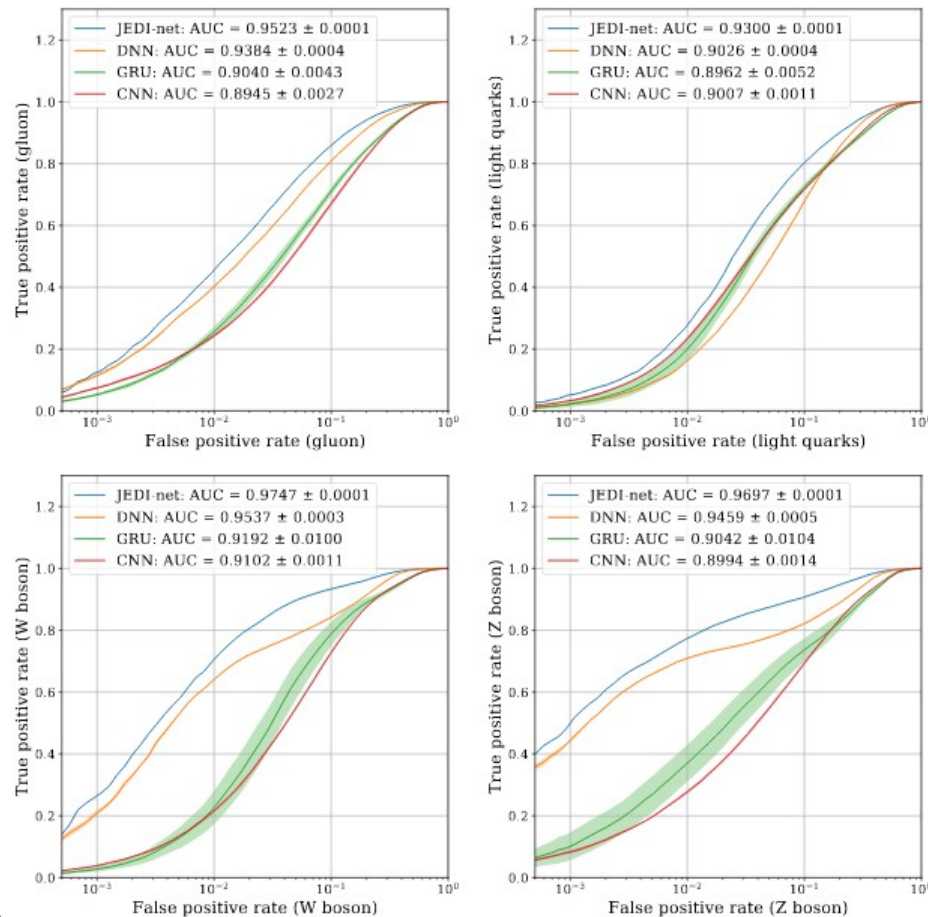
Graph-level classification (binary or multi-class)



Classification Performance

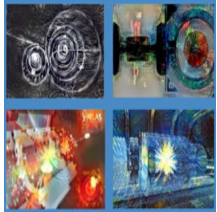


Out-perform other deep learning methods for jet multi-class categorization





Remarks

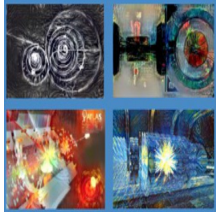


- *Single step of graph network thanks to all-to-all connectivity*
- *Extraction of a graph-level feature can be done in multiple ways.*
- *De-correlation with learning the mass done with adversarial, DDT, ...*
- *What have we done with the Physics knowledge of jet metrics ?*

<https://github.com/eric-moreno/IN>



Summary and Outlooks



- Graphs are a very natural data representation in HEP.
 - Deep learning on graph helps on several HEP tasks.
 - Multiple ways of doing deep machine learning on a graph.
 - Further application of graph network to HEP to come.

Get in touch !
jvlimant@caltech.edu